



# UNITED STATES AIR FORCE RESEARCH LABORATORY

---

## Head-Related Transfer Functions in the Near Field

Douglas S. Brungart

AIR FORCE RESEARCH LABORATORY

William M. Rabinowitz

BOSE CORPORATION  
Boston MA

March 1998

Interim Report for the Period September 1995 to March 1998

20020307 054

*Approved for public release; distribution is unlimited.*

Human Effectiveness Directorate  
Crew System Interface Division  
2610 Seventh Street  
Wright-Patterson AFB OH 45433-7901

## NOTICES

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Please do not request copies of this report from the Air Force Research Laboratory. Additional copies may be purchased from:

National Technical Information Service  
5285 Port Royal Road  
Springfield, Virginia 22161

Federal Government agencies and their contractors registered with the Defense Technical Information Center should direct requests for copies of this report to:

Defense Technical Information Center  
8725 John J. Kingman Road, Suite 0944  
Ft. Belvoir, Virginia 22060-6218

## TECHNICAL REVIEW AND APPROVAL

AFRL-HE-WP-TR-2001-0171

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public.

This technical report has been reviewed and is approved for publication.

**FOR THE COMMANDER**



MARIS M. VIKMANIS  
Chief, Crew System Interface Division  
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE March 1998		3. REPORT TYPE AND DATES COVERED Interim - September 1995 to March 1998
4. TITLE AND SUBTITLE  Head-Related Transfer Functions in the Near Field			5. FUNDING NUMBERS  PE - 62202F PR - 7184 TA - 718441 WU - 71844104	
6. AUTHOR(S) Douglas S. Brungart, AFRL William M. Rabinowitz, Bose Corp.				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Air Force Research Laboratory, Human Effectiveness Directorate Crew System Interface Division Bioacoustics and Biocommunications Branch Air Force Materiel Command Wright-Patterson AFB OH 45433-7901			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSORING/MONITORING AGENCY REPORT NUMBER  AFRL-HE-WP-TR-2001-0171	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION AVAILABILITY STATEMENT  Approved for public release; distribution is unlimited.			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words)  Although researchers have long recognized the unique properties of head-related transfer functions (HRTF) in the near field (within 1 m of the listener's head), virtually all of the HRTF measurements described in the literature have focused on source locations 1 m or farther from the listener. A rigid-sphere model of the head has been used to calculate the near-field HRTF, and the HRTF was measured at distances of 0.12 m, 0.25 m, 0.50 m, and 1.0 m with a KEMAR manikin and an acoustic point source. Both the calculations and the measurements indicate that the interaural intensity difference increases substantially for lateral sources as distance decreases below 1 m, while the interaural time delay remains roughly independent of distance. The high-frequency features of the KEMAR HRTFs were compressed in azimuth around the ipsilateral ear when the source was close. The elevation-dependent features of the HRTFs were not strongly dependent on distance, and the contribution of the pinna to the HRTF was independent of distance beyond a few centimeters from the ear. The measurements indicate that binaural source-distance cues, which occur only with nearby sources, may allow listeners to make judgments about the distances of nearby sound sources.				
14. SUBJECT TERMS  head-related transfer functions, auditory localization, virtual audio displays			15. NUMBER OF PAGES 71	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT  UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE  UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT  UNCLASSIFIED	20. LIMITATION OF ABSTRACT  UNL	

THIS PAGE INTENTIONALLY LEFT BLANK

## PREFACE

The work described herein was performed at the Noise and Vibration Branch, Force Survivability and Logistics Division, Human Effectiveness Directorate of the Air Force Research Laboratory and the Research Laboratory of Electronics at the Massachusetts Institute of Technology. At the time the research was performed, Douglas S. Brungart was a participant in the Palace Knight civilian training program at MIT, and the work described was incorporated in his doctoral dissertation (Brungart, 1998). William Rabinowitz was at that time a Principle Research Scientist at the Sensory Communication group of the Research Laboratory of Electronics and was serving on Dr. Brungart's graduate committee. The authors would especially like to thank Bill Peake for his helpful comments about the figures, and Nat Durlach and Steve Colburn for their suggestions about the manuscript. They would also like to thank Dennis Allen for his assistance in setting up the HRTF measurements.

THIS PAGE INTENTIONALLY LEFT BLANK

# TABLE OF CONTENTS

Preface	iii
List of Figures	vii
1.0 Introduction	1
2.0 Background	2
3.0 The HRTF in the near field	5
4.0 Sphere model of the head	9
5.0 Procedures for HRTF measurements	11
5.1 Facilities . . . . .	11
5.2 Measurement Procedure . . . . .	12
5.3 Calibration . . . . .	13
6.0 Monaural transfer functions	15
6.1 Evaluation of Monaural Transfer Functions with Sphere Model . . . . .	15
6.1.1 Head shadowing . . . . .	17
6.1.2 Source proximity . . . . .	17
6.1.3 High-frequency pressure doubling . . . . .	19
6.1.4 Acoustic bright spot . . . . .	19
6.1.5 Low-frequency diffraction . . . . .	20
6.1.6 Low-pass filtering as a possible spectral distance cue . . . . .	20
6.1.7 Contour plots of sphere-model HRTFs . . . . .	20
6.2 Measurements of Monaural Transfer Functions with the KEMAR Manikin	23
6.2.1 Comparison of KEMAR measurements with sphere model . . . . .	23
6.2.2 Distance dependent changes in the high-frequency ipsilateral HRTF .	26
7.0 Interaural intensity differences	29
8.0 Interaural time delays	31
9.0 Elevation effects on near-field HRTFs	34

10.0 Pinnae effects for very near sources	36
11.0 Comparison of sphere model and KEMAR measurements	39
12.0 Perceptual implications of the near-field HRTFs	41
13.0 Conclusions	45
14.0 References	47
<b>Appendix A: Acoustic point source for near-field HRTF measurements</b>	<b>51</b>
A.1 Abstract . . . . .	51
A.2 Purpose . . . . .	51
A.3 Background . . . . .	52
A.4 Description of System . . . . .	56
A.5 Characteristics of Point Source . . . . .	58
A.6 Use of the Point Source . . . . .	61

## LIST OF FIGURES

FIGURE	PAGE
1 The near and far fields for auditory localization . . . . .	6
2 Sphere model HRTFs in horizontal plane . . . . .	16
3 Head shadowing . . . . .	18
4 Low-pass filtering with distance . . . . .	21
5 Contour plots of monaural HRTF . . . . .	22
6 KEMAR HRTFs in horizontal plane . . . . .	24
7 High-frequency HRTF distance effects . . . . .	27
8 Interaural intensity differences . . . . .	30
9 Interaural time delays . . . . .	32
10 Distance/elevation interactions in HRTF . . . . .	35
11 Near-field pinna effects . . . . .	37
12 Comparison of sphere model and KEMAR measurements . . . . .	40
A-1 Diagram of acoustic point source . . . . .	57
A-2 Source transfer function . . . . .	59
A-3 Source directionality . . . . .	59
A-4 Source response vs. distance in anechoic chamber . . . . .	60
A-5 Source linearity . . . . .	61
A-6 Flattened pink noise output from point source . . . . .	63

THIS PAGE INTENTIONALLY LEFT BLANK

## 1.0 INTRODUCTION

Over the past 100 years, considerable efforts have been made to characterize the relationship between the location of a sound source in space and the sound pressure generated by that source at the eardrums of a human listener. This relationship has generally been described in the frequency domain as the head-related transfer function, or HRTF. Systematic variations in the HRTF with azimuth and elevation have been studied extensively and are well documented. These results indicate that the HRTF is roughly independent of distance when the source is more than 1 m away from the head. Despite the recognition by earlier researchers that the HRTF varies significantly with distance at distances less than 1 m (Stewart, 1911; Hartley & Frey, 1921; Firestone, 1930; Blauert, 1983), the dependence of the HRTF on distance remains largely unexplored. This paper examines the behavior of the head-related transfer function at distances less than 1 m. The first two sections provide background on HRTF measurements and the changes that occur in the HRTF as the source approaches the head. These sections are followed by a description of the rigid-sphere model and of the procedures used to make the acoustic measurements. The next five sections show the results of the model and KEMAR measurements in terms of the monaural HRTF in the horizontal plane, the interaural intensity difference, the interaural time delay, the interaction between distance and elevation in the monaural HRTF, and the effects of the pinna in the near field. Finally, the perceptual implications of the distance-dependent features of the HRTF are briefly discussed.

## 2.0 BACKGROUND

The ability of human listeners to identify the location of a sound source has been studied extensively in the past century. Lord Rayleigh (1907) first observed that sound localization is made possible by the geometric properties of the human head and, in particular, by the location of the ears at opposite sides of the head. This ear placement results in two important binaural cues for localizing the azimuth of a sound source: a delay between the arrival of the sound at the left and right ears, known as interaural time delay (ITD), and a difference in pressure level at the near and far ears, known as interaural intensity difference (IID). Rayleigh reasoned that the interaural time delay provided an unambiguous localization cue only when the ITD was less than one full period of the sound wave, i.e., at frequencies below approximately 2 kHz. Similarly, the IID is a useful cue only at high frequencies (above 3 kHz) where there is significant head-shadowing at the contralateral ear. The notion that the ITD is the primary localization cue at low frequencies and IID is the primary localization cue at high frequencies is known as the Duplex Theory.

Later researchers recognized some important limitations in the Duplex Theory. Interaural differences, on which the theory is based, do not correspond to a unique source location. To a first order approximation, both the ITD and IID are determined only by the angle between the interaural axis and the sound source. Consequently, ITD and IID information cannot distinguish between source locations on the surface of a cone centered on the interaural axis with its apex at the center of the head, a locus of points known as the “cone of confusion” (Wallach, 1939). Yet human listeners typically do not make errors along the cone of confusion during auditory localization, at least for wideband stimuli. Directionally dependent changes in the sound spectrum reaching the ear, produced both by the diffraction of sound by the head and torso and by the intricate shape of the outer ear or pinna, allow listeners to make accurate front-back and up-down judgments about sound location (Musicant & Butler, 1984; Oldfield & Parker, 1984). [For a thorough review of auditory localization cues, see Middlebrooks & Green (1991).]

The relationship between the sound originating from a point source in space and the sound actually reaching the eardrum of a listener is expressed by the head-related transfer function, or HRTF. The HRTF includes both magnitude and phase information as well as

the effects of the locations of the ears, diffraction by the head and torso, spectral shaping by the pinnae, and the resonance of the ear canal. The precise definition of the HRTF varies. Any direct measurement typically includes the frequency response of the loudspeaker generating the stimulus and the microphone used to make the measurement, and the desired transfer function from the source to the eardrum. In the literature, measured HRTFs have been presented in a variety of ways, including:

1. The ratio of the output of a probe microphone location 1-2 mm from the eardrum of a human subject to the input of the loudspeaker (Wightman & Kistler, 1989).
2. The ratio of the output of a probe microphone near the eardrum to the free-field pressure at the location of the probe microphone with the head removed (Pralong & Carlile, 1994; Mehrgardt & Mellert, 1977).
3. The ratio of the sound pressure at the opening of a blocked ear canal to the free-field sound pressure at the center of the head with the head removed (Moller, Sørensen, Hammershoi, & Jensen, 1995).
4. The ratio of the sound pressure at the eardrum to the free-field sound pressure at the center of the head with the head removed (Gardner & Martin, 1995).
5. The ratio of the sound pressure in the ear canal to the sound pressure in the canal with the source directly in front of the listener (Shaw, 1974).
6. The ratio of the sound pressure in the left ear canal to the sound pressure in the right ear canal (the interaural HRTF) (Firestone, 1930; Carlile & Pralong, 1994).
7. The ratio of the sound pressure in the ear canal to the maximum sound pressure (over all locations) at that frequency (used to show the directionality of the HRTF) (Middlebrooks, Makous, & Green, 1989).

These are just a few of the definitions of the HRTF which have been used in the literature. Although the actual location of the microphone within the ear canal has a strong effect on the measured HRTF, this effect is largely independent of the sound location (Middlebrooks et al., 1989). In other words, the HRTFs measured at different locations in the ear canal will vary only by a linear transformation which is independent of direction. The relative changes in the HRTF with direction are preserved regardless of the position of the microphone within the ear canal. It should be recognized, however, that the sound pressure measured in the

ear canal is not equivalent to the sound pressure at the eardrum, and some investigators have attempted to estimate the transformation from ear canal to eardrum to evaluate the true source-to-eardrum HRTF (Chan & Geisler, 1990; Moller et al., 1995). Note that the last three HRTF definitions are defined only relative to the HRTF at some other location and, therefore, they do not reflect any direction-independent features of the HRTF. In the measurements described in this paper, the “monaural” HRTF is defined as the ratio of the sound pressure at the eardrum to the free-field sound pressure at the location of the center of the head.

Another major difference among HRTF measurements in the literature is the use of humans and manikins in the HRTF measurements. Several studies have used acoustic manikins (Kuhn, 1979; Firestone, 1930; Gardner & Martin, 1995), which have some important advantages: they allow microphone placement at the exact location of the eardrum, do not require immobilization during measurements, and can remain in place indefinitely during a long series of measurements. The majority of studies, however, have used human subjects, both because they eliminate uncertainty about the similarity between human and manikin ears and because they allow comparison of the HRTFs across a population of subjects. In the past decade, technological advances including automated source placement systems (Wightman & Kistler, 1989; Middlebrooks et al., 1989), and fast, high signal-to-noise ratio measurements (Foster, 1986) have allowed researchers to measure HRTFs on human subjects more quickly. Furthermore, much research in recent years has focused on the use of HRTFs in virtual audio displays (see Wenzel (1991) for review), and on the importance of the detailed features of one’s own HRTFs in producing realistic virtual sounds (Wightman & Kistler, 1989). Each of these developments has increased the emphasis on making HRTF measurements with human subjects.

Despite the wide variety of procedures used in previous HRTF measurements, there has been no serious effort to study the effects of distance on the HRTF. With the exception of the early study by Firestone (1930), nearly all of the HRTF measurements presented in the literature were made with sound sources located 1 m or further from the listener. The importance of source distance has not been emphasized in these measurements because the HRTF is roughly independent of distance in this region. The next section briefly discusses the unique properties of the near field in acoustics, and the changes that are expected to occur in the HRTF at distances less than 1 m.

### 3.0 THE HRTF IN THE NEAR FIELD

In acoustics, the definitions of the terms “near field” and “far field” depend on context. The distance-dependent changes in the auditory localization cues occur when the source approaches within one meter of the head. Therefore, in terms of human sound localization, we will designate the near field as the region of space within 1 m of the center of the listener’s head and the far field as the region at distances greater than 1 m.

The fundamental differences between the near field and the far field are illustrated in Figure 1. If diffraction by the head is ignored, there are two primary differences between the HRTF in the near and far fields. First, the decrease in the intensity of the radiating sound wave with distance (illustrated by the decreasing thickness of the lines) is large over the region occupied by the head in the near field, but small over the region occupied by the head in the far field. In this example, the intensity decreases by 10 dB from the nearest to the furthest point on the closer head, but only by 1.75 dB on the more distant head. As a result of this intensity effect, the amplitude of the sound at the ipsilateral ear increases more rapidly than the amplitude at the contralateral ear as a near-field source approaches the head. Therefore the interaural intensity difference is larger for nearby sources than for distant sources. In contrast, at distances greater than 1 m, the intensity of the sound wave is not significantly different at the locations of the ipsilateral and contralateral ears, and the IID is essentially independent of distance.

Second, the orientation of each point on the surface of the head relative to the point source varies significantly for the nearby source, but is roughly constant for the more distant source. In this figure, the angle from the nose to the source differs from the angle from the ipsilateral ear to the source by approximately  $50^\circ$  for the nearby source, but only by  $9^\circ$  for the more distant source. Since diffraction by the head depends, in part, on the angle of incidence of the sound wave impinging on the surface of the head, changes in the orientation of the source over the surface of the head can significantly influence the near-field HRTF.

Note that the interaural time delay, which depends on the absolute propagation delay between the ipsilateral and contralateral ears and not on the ratio of the distances between left and right ears and the source, is much less dependent on distance in the near field than the IID.

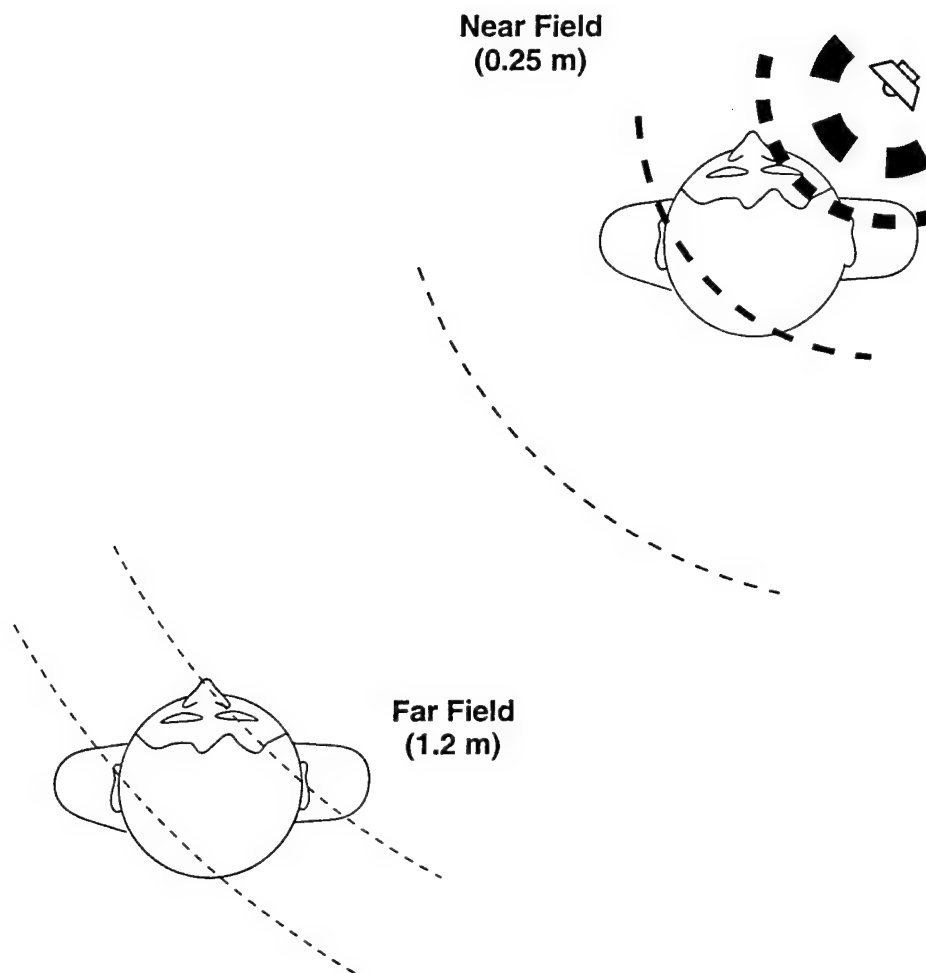


Figure 1: Comparison of near-field and far-field localization cues. The intensity of the spherically radiating sound wave is indicated by the thickness of the lines.

There are some practical issues which make the measurement of HRTFs in the near-field relatively difficult. The most important issue is the size of the sound source used to make the measurement. Traditionally, HRTF measurements have used conventional loudspeakers, 7 cm or larger in diameter, to generate a free-field signal. At distances of 1 m or more, loudspeakers are perfectly adequate. At close distances, however, there are serious problems associated with loudspeaker measurements:

- The precise location of a loudspeaker is not well defined in the near-field. The stimulus is generated by the entire diaphragm of the loudspeaker, and at close distances this may extend over a large region of space: at 12 cm, for example, a 7 cm loudspeaker covers an arc in excess of 30°. The HRTF measured will be, in effect, the average HRTF over the entire region covered by the loudspeaker.
- The directional properties of the loudspeaker may taint the HRTF. When the speaker is near the listener, the high-frequency directionality of the speaker will cause the sound pressure reaching the head and torso to vary according to the orientation of that region relative to the speaker. This may significantly affect the measured HRTF.
- The axial response of a loudspeaker is complicated by its distributed geometry at very close distances. At distances less than  $2\frac{a^2}{\lambda}$ , where  $a$  is the radius of the loudspeaker and  $\lambda$  is the wavelength of the sound, the intensity along the axis of the loudspeaker does not decrease monotonically with distance, but rather passes through a series of maxima of constant amplitude with intervening nulls (Kinsler & Frey, 62). For a 15 kHz sound generated by a 7 cm loudspeaker, this effect complicates measurements at distances less than 10 cm from the surface of the head (approximately 20 cm from the center of the head).
- A large loudspeaker may reflect the sound diffracted by the sphere, generating a standing wave between the speaker and the head and corrupting the HRTF measurements.

For these reasons, a loudspeaker cannot be used effectively to make near-field HRTF measurements. Therefore an approximation to an acoustic point source has been developed for this set of experiments, as described in the measurement procedure section.

Accurate placement of the head during the HRTF measurement is also more difficult in the near-field. An error in placement of a few centimeters is irrelevant for a source 1 m away from the head, but critical for a nearby source. Finally, the automated source placement

systems widely used to make HRTF measurements are not easily adaptable to three dimensions. These difficulties may explain the absence of near-field HRTF measurements from the recent literature.

## 4.0 SPHERE MODEL OF THE HEAD

Approximate interaural differences associated with a nearby sound source can be obtained from mathematical descriptions of the acoustic properties of rigid spheres. This approach to modeling near-field HRTFs is not new, and in fact was used to make manual calculations about near-field IIDs by Stewart (1911). Hartley and Frey (1921) manually tabulated interaural amplitudes and time delays at a variety of distances and directions using Stewart's derivation. The model described in this section was adapted from the work of Rabinowitz et al. (1993), who examined the frequency scalability of head-related transfer functions for an enlarged head. This variation of their model maintains a fixed head size and varies distance, rather than varying head size at a fixed distance (Brungart & Rabinowitz, 1996). Duda (1997) has compared the predictions of this model to acoustic measurements made on the surface of a bowling ball.

The model approximates the head as a rigid sphere, of radius  $a$ , with "ears" located at diametrically opposed points on the surface of the head. The sound source is a point velocity source radiating spherical acoustic waves, and is located at distance  $r$  from the center of the head and at angle  $\alpha$  from the perpendicular bisector of the interaural axis. The complex expression for the sound pressure at the ear, denoted by  $P_s$ , is given by

$$P_s(r, a, \alpha, f) = \frac{c\rho_0 u_0}{2\pi a^2} \sum_{m=0}^{\infty} \left(m + \frac{1}{2}\right) L_m(\cos \alpha) \frac{H_m\left(\frac{2\pi fr}{c}\right)}{H'_m\left(\frac{2\pi fa}{c}\right)} e^{-j\frac{\pi}{2}}, \quad (1)$$

where  $a$  is the radius of the sphere,  $f$  is the sound frequency (in Hz),  $c$  is the speed of sound,  $u_0$  is the volume velocity of the (infinitesimal) source,  $L_m$  is the Legendre polynomial function, and  $H_m$  is the spherical Hankel function.

In order to calculate the monaural transfer function from the model, the pressure at the ear must be divided by the reference pressure at the center of the head, which is simply the output of a point source of strength  $u_0$  at distance  $r$ , or

$$P_{ff}(r, f) = \frac{2\pi f \rho_0}{u_0} 4\pi r e^{j\left(-\frac{2\pi fr}{c} + \frac{\pi}{2}\right)}. \quad (2)$$

A remaining complication is the transformation of the angle  $\alpha$  between the location of the source and the location of the ear to the more standard spherical coordinates used to define HRTFs. Throughout this paper, the coordinate system has its origin at the midpoint of the interaural axis. Azimuth ( $\theta$ ) will be defined as  $0^\circ$  directly in front of the head,  $90^\circ$  directly to the left, and  $-90^\circ$  directly to the right. Elevation ( $\phi$ ) will be  $0^\circ$  in the horizontal plane,  $90^\circ$  directly above, and  $-90^\circ$  directly below. In this coordinate system, the monaural transfer function at the left ear is

$$\mathbf{H}_L(r, a, \theta, \phi, f) = \frac{\mathbf{P}_s(r, a, \arccos(\sin(\theta) \cos(\phi)), f)}{\mathbf{P}_{ff}(r, f)}. \quad (3)$$

## 5.0 PROCEDURES FOR HRTF MEASUREMENTS

### 5.1 Facilities

The HRTF measurements were made inside the anechoic chamber of the Armstrong Laboratory at Wright-Patterson Air Force Base, a large chamber (10 m x 10 m x 10 m) which currently contains the Auditory Localization Facility (ALF). The ALF is a large, wire-frame geodesic sphere used in localization experiments at the Armstrong Laboratory (McKinley, Ericson, & D'Angelo, 1994). Because of the ALF, it was necessary to make the HRTF measurements in a corner of the anechoic chamber. Acoustic measurements with a free-field microphone verified that the presence of the ALF did not significantly impair the anechoic conditions in the corner where the measurements were made for source distances up to 1 m. All of the HRTF measurements were made with a Knowles Electronic Manikin for Acoustic Research (KEMAR). The KEMAR manikin consists of an anthropomorphic rigid plastic head and torso. The left and right pinnae are constructed of soft rubber and mounted in removable panels on the sides of the manikin head. Inside each manikin ear, a Zwislocki coupler simulates the acoustic properties of the ear canal and the middle ear impedance, and a Bruel & Kjaer 1.2 cm (0.5") pressure microphone attached to the coupler measures a pressure approximately equivalent to that at the eardrum of a human listener. The output of the left and right microphones was connected to a Bruel & Kjaer 5935 dual microphone power supply, and then passed through a patch panel into the control room.

The KEMAR was mounted on a metal stand equipped with optically-encoded stepper motors which allow electronic control of the azimuth and elevation of the manikin within a fraction of a degree.

The sound source used in the measurements was an approximation to a point source. A sound driver was connected to a 3 m long piece of Tygon tubing with an internal diameter of 1.2 cm and 1.5 mm thick walls. For convenience, the end of this tube was mounted in a PVC pipe sleeve, 2.5 cm in diameter and 64 cm in length, with the end of the tube projecting 2 cm from the end of the pipe and foam material sealing the space between the tube and the interior of the sleeve. The sleeve was used to clamp the point source to a tripod stand which was used to position the source during the measurements.

The relatively small opening of the tube allows this source to act approximately as an acoustic point source in that it is essentially non-directional even at high frequencies. At 15 kHz, for example, the 3-dB beam-width of the source was found to be 120°. Furthermore, the small size allows a precise placement of the source and eliminates the potential problem of secondary reflections off the source. Although the frequency response of this source is highly irregular, its shape is consistent and easily removed from the transfer function measurements, and its effective frequency range is from 200 Hz to 15 kHz.

In order to calculate the HRTF, reference measurements were made with a Bruel & Kjaer 1.2 cm (0.5") free-field microphone located at the position of the center of the manikin head with the manikin removed. Measurements were made at 0.125 m, 0.15 m, 0.25 m, 0.50 m, and 1.00 m before and after the HRTF measurements. Changes in the frequency response of the source over the course of the measurements were found to be negligible (within  $\pm 1.5$  dB) and the signal measured at the response microphone was essentially independent of distance except for the inverse relation of overall amplitude to distance.

The measurements were controlled by a computer located in a small room adjacent to the anechoic chamber. The two rooms were connected by a patch panel which passed the left and right microphone signals, the sound source signal, and the motor controller signal. The computer was also connected via GPIB bus to a Hewlett-Packard HP35665A dynamic signal analyzer which was used both to generate the source signal and to measure the transfer functions. The microphone signals were connected to the analyzer inputs. The source signal was amplified by a Crown D-75 Power Amp before being passed through the patch panel to the sound driver.

## 5.2 Measurement Procedure

In all of the measurements, the HP35665A was operated in transfer-function mode, which measures the ratio of the power spectrum on the second channel to the power spectrum on the first channel. In this mode, one of the ear microphones is connected to the second channel, and the first channel is connected directly to the source output of the analyzer. A periodic chirp source signal was used, in conjunction with a uniform window, to maximize the signal-to-noise ratio. At each source position, 64 FFT measurements were averaged using RMS averaging.

All measurements were made at two frequency ranges: from 100 Hz to 12.9 kHz and from 7.78 kHz to 20.6 kHz. Each measurement consisted of a 400-point FFT with 32 Hz

resolution. The measurements were then combined, with the first 240 points of the low-frequency measurement and the first 360 points of the high-frequency measurement, to give an overall 600-point transfer function with 32 Hz resolution from 100 Hz to 19.2 kHz. The two-part measurement was used for two reasons. First, the measurement allowed higher resolution over the entire range of human hearing than a single 400-point measurement with 64 Hz resolution. Second, and more importantly, the dual measurements allowed independent ranging of the HP35665A input channel at low and high frequencies. The transfer function of the point source has an approximately 20 dB drop-off in frequency response around 7.5 kHz, and in a single measurement the analyzer either overloaded at low frequencies or approached the noise floor at high frequencies. By dividing the measurement into two parts, it was possible to adjust the analyzer to maximize the signal-to-noise ratio in both frequency bands without overloading. Proper ranging in each frequency band was ensured by the control computer, which forced the signal analyzer to auto-range prior to each measurement in each frequency band, and repeated any measurements where an overload occurred with a slightly higher input range.

After each measurement, the amplitude and phase of the transfer function at each frequency value were saved into separate ASCII files. The monaural HRTF at each location was found by dividing the amplitude spectrum at that location by the calibration measurement corresponding to the source distance. The phase files were used to calculate interaural time delays. First, the phase spectrum at the right ear was subtracted from the phase spectrum at the left ear for a given source direction and distance. Then the time delay was calculated by finding the constant time delay closest to the measured phase difference (in the least-squared-error sense) in the frequency range of 100 Hz to 6500 Hz. Note that the squared error is based on the angular error which is restricted to the range  $-180^\circ$  to  $180^\circ$ . The calculated time delay may also be viewed as the slope of the line which best fits the unwrapped phase difference between the left and right ears. This time delay measurement has been found to be repeatable within 1-2  $\mu s$ .

### 5.3 Calibration

Prior to each set of measurements, the KEMAR manikin was carefully positioned to place the center of the interaural axis directly over the axis of rotation of the stand. In near-field measurements, correct placement is particularly important because even a small deviation between the center of the head and the center of rotation will cause the distance

from the source to the center of the head to change as a function of azimuth and severely corrupt the transfer function measurements at very close distances. The centering of the KEMAR head was accomplished automatically with a series of acoustic measurements with the source placed in front of the manikin at  $0^\circ$  elevation. First, the manikin was centered in azimuth by rotating the head until the magnitude of the interaural time delay was reduced below  $2 \mu s$ . Then the roll of the manikin was verified by rotating the KEMAR  $180^\circ$  and again verifying that the time delay was approximately  $0 \mu s$ . With a nearby source, any left or right tilt of the manikin will prevent the source from falling on the median plane at both  $0^\circ$  and  $180^\circ$  in azimuth. Finally, the manikin head was centered in elevation by verifying that the low-frequency time delay from the source to the left ear at  $0^\circ$  in azimuth was equivalent (within  $5 \mu s$ ) to the delay from the source to the left ear at  $180^\circ$ . If the manikin were tilted forward or backward, the distance (and therefore delay) from the source to the ear would be different when the manikin was facing  $0^\circ$  than when facing  $180^\circ$ . After the elevation was adjusted, the centering measurements were repeated until the manikin was acceptably centered in azimuth, elevation, and roll. Note that, while the adjustments in azimuth and elevation were completely automated, the adjustments in roll were made manually by inserting material between the base of the manikin and the motorized stand. This was not a serious limitation, however, because yaw only required adjustment once prior to the measurement procedure.

## 6.0 MONAURAL TRANSFER FUNCTIONS

### 6.1 Evaluation of Monaural Transfer Functions with Sphere Model

First consider the monaural transfer function of the left ear predicted by the sphere-based mathematical model of the head. The important characteristics of these transfer functions (Figure 2) can be summarized by four observations:

1. The magnitude of the HRTF increases across all frequencies as the ear rotates toward the source, and decreases as the ear rotates away from the source.
2. The magnitude of the HRTF increases with frequency when there is a direct path between the sound source and the ear, and decreases with frequency when the ear lies in the acoustic shadow of the head.
3. The magnitude of the HRTF decreases with distance when there is a direct path from the source to the ear, and increases with distance when the ear is shadowed by the head.
4. The monaural HRTFs change rapidly as distance decreases below 0.5 m, but change by no more than 1 dB as distance increases from 1 m to 10 m.

Many of the features of the sphere-model HRTFs can be explained intuitively with relatively simple acoustic concepts. Head shadowing effects are primarily responsible for low-pass filtering the signal at the contralateral ear. Source proximity effects contribute to the change in the magnitude of the transfer function with azimuth at low frequencies, and to the changes in the magnitude of the transfer function with distance. High-frequency pressure doubling causes the magnitude of the ipsilateral ear transfer functions to increase with frequency. And the acoustic "bright-spot" causes the attenuation at high frequencies to decrease when the ear is pointing directly away from the source. Each of these concepts is explained in more detail below.

## HRTFs for Left Ear Predicted by Sphere Model

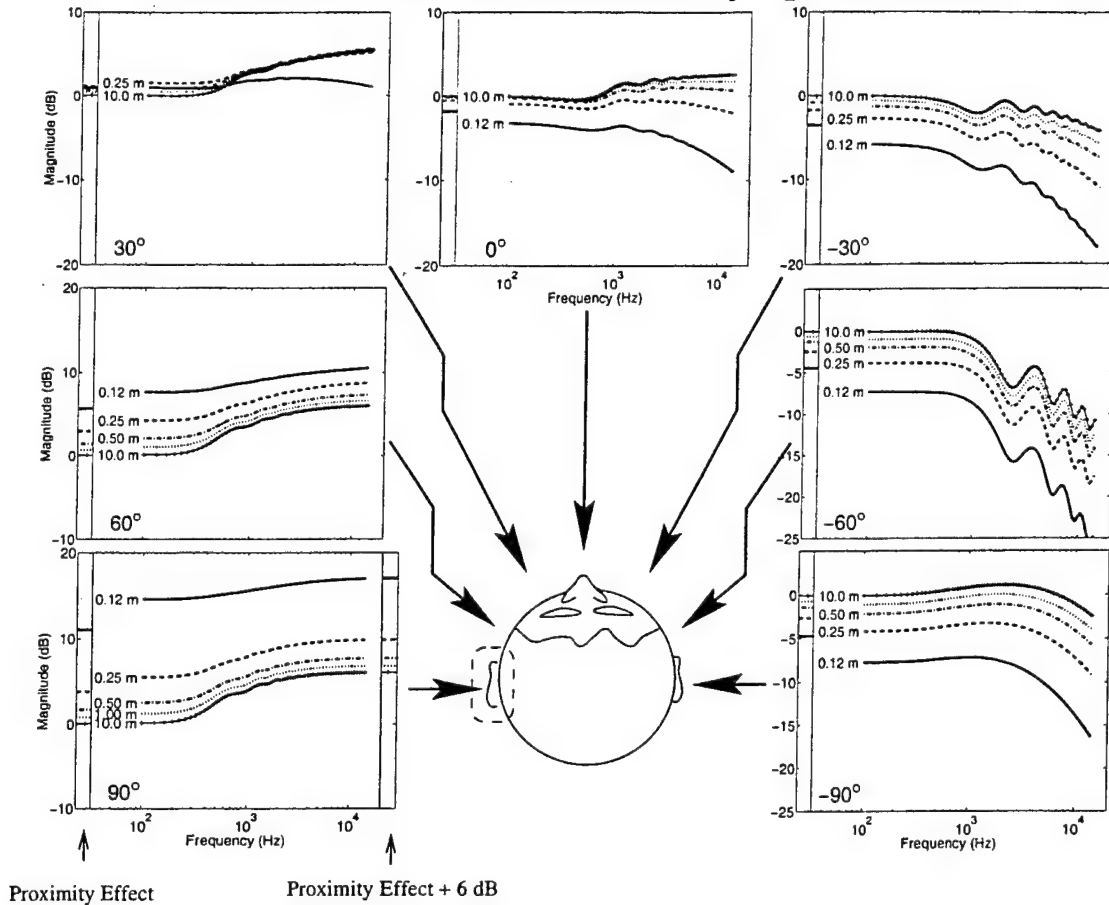


Figure 2: The monaural HRTF for horizontal-plane sources from 0.125 m to 10 m when the head is a rigid sphere 18 cm in diameter. The HRTFs were calculated by dividing the pressure at the left ear by the free-field pressure at the center of the head (see Equation 3). Results are provided at 30° intervals in the front hemisphere only, as the sphere model is symmetric across the frontal plane. Frequency is shown at 100 Hz intervals from 100 Hz to 1 kHz, and at 1/12 octave intervals from 100 Hz to 15 kHz. The bars at the left side of each graph show the source proximity effect, which is the gain of the HRTF ignoring diffraction by the head. In the 90° graph, the bar at the right side of the figure shows the source proximity effect plus the 6 dB high-frequency pressure doubling effect, and illustrates that the combination of these two effects fully explains the high-frequency asymptotes of the HRTFs at this location. See text for details.

### 6.1.1 Head shadowing

Head shadowing is simply the attenuation in the HRTF that occurs when the head obscures the direct path from the sound source to the ear. The effect of the shadow is related to the size of the head relative to the wavelength of the sound, so the attenuation at the shadowed ear increases with frequency. As a result, the HRTFs for the contralateral ear resemble low-pass filters (see the HRTFs for  $-30^\circ$ ,  $-60^\circ$ , and  $-90^\circ$  in Figure 2).

As the source approaches the head, both the size and attenuation of the shadowed zone increase. The size increases because of the convexity of the spherical head. No unoccluded path exists from a point on a convex surface to the region inside the plane tangent to the surface at that point. In the case of the spherical head, the left ear is shadowed for all sources located to the right of the plane tangent to the head at the left ear (Figure 3). This region includes  $0^\circ$  azimuth at all distances, and  $30^\circ$  azimuth when the source is closer than 18 cm. Thus, at  $30^\circ$  in Figure 2, the ear is in the acoustic shadow of the head only at 0.12 m, and the high-frequency response of the HRTF at 0.12 m is attenuated by the head shadow.

The amount of attenuation due to the head shadow increases as the ear is located further inside the shadowed region. Since the size of the shadowed region increases as source distance decreases, this results in increased high-frequency attenuation for nearby sources at the shadowed ear. The increase in high-frequency attenuation with decreasing source distance is seen at  $0^\circ$ ,  $-30^\circ$ , and  $-60^\circ$  in Figure 2.

### 6.1.2 Source proximity

The inverse relationship between pressure and distance for a spherically radiating sound wave can significantly influence the HRTF when the source is near the head. This *source-proximity* effect can be viewed as the portion of the HRTF which is not a result of diffraction by the head, i.e., as the HRTF for a pressure sensor suspended in free space at the location of the ear. Since the HRTF is defined as the ratio of the pressure at the ear to the pressure at the center of the head, the source-proximity effect is simply the ratio of the distance from the source to the center of the head to the distance from the source to the ear.

The magnitude of the source-proximity effect is shown along the left side of each panel in Figure 2. At 10.0 m, the effect is negligible (near 0 dB) for all source directions. The effect increases as source distance decreases, and at 0.12 m the source-proximity effect produces

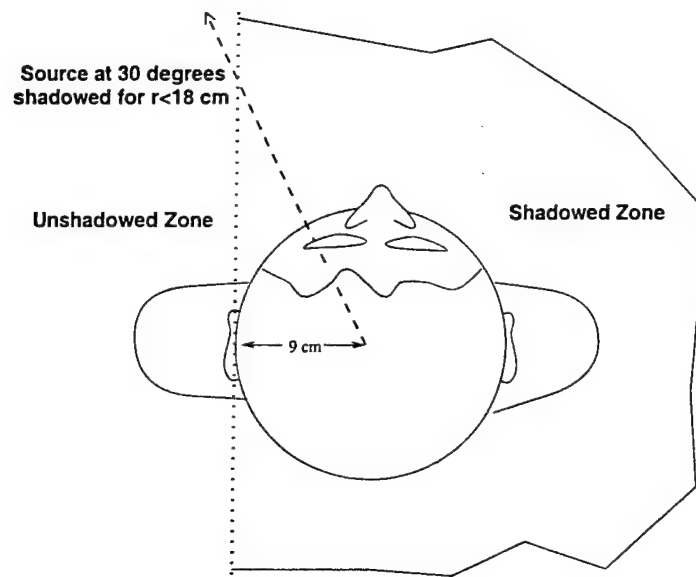


Figure 3: Head shadowing of a source at the ipsilateral ear. The boundary between the shadowed and non-shadowed regions for the ear is plane tangent to the sphere at the location of the ear. In this illustration of the shadowed region for the left ear, sources in the shaded region are shadowed and sources in the unshaded region have a direct line of sight to the ear. The line shows a source at  $30^\circ$ , where a source is shadowed when it is closer than 18 cm to the center of the head. Note that sources at all distances are shadowed when directly in front of the listener.

more than 10 dB of gain at  $90^\circ$  and almost 5 dB of attenuation at  $-90^\circ$ . The source-proximity effect can explain some, but not all, of the low-frequency behavior of the HRTF. In general, the ordering of the low-frequency HRTFs by distance is consistent with the source-proximity effect at each azimuth location, but the magnitude of the low-frequency gain or attenuation is greater than that predicted by source proximity.

### **6.1.3 High-frequency pressure doubling**

The magnitude of the HRTF at the ipsilateral ear generally increases with frequency due to high-frequency reflections off the surface of the sphere. When the source is located at  $90^\circ$ , the sound waves impinging on the ear are perpendicular to the surface of the sphere, and at sufficiently high frequencies the sound wave is specularly reflected off the surface of the sphere back in the direction of the source. At the surface of the sphere, the direct and reflected sound waves combine in phase to produce a 6 dB pressure gain. This 6 dB increase in high-frequency gain is evident in the 10 m HRTF at  $90^\circ$  in Figure 2. In fact, the high-frequency magnitude of the HRTF at  $90^\circ$  is exactly equivalent to the combination of the source proximity effect and the high-frequency pressure doubling effect, as shown at the right side of the panel in Figure 2. As the source rotates away from  $90^\circ$ , the sound waves from the source are no longer perpendicular to the surface of the sphere and only a portion of the sound wave is reflected at the surface, resulting in a high-frequency gain less than 6 dB at  $60^\circ$  and  $30^\circ$ . Note that high-frequency pressure doubling does not occur in the contralateral HRTFs.

### **6.1.4 Acoustic bright spot**

When the ear is located directly opposite the source ( $-90^\circ$  in Figure 2), all of the possible sound paths from the source to the ear are cylindrically symmetric and, consequently, all of the components of the diffracted sound wave combine in phase at the ear. This in-phase combination results in a local maximum in the HRTF at all frequencies. The resulting phenomenon, known as the acoustic "bright spot," is clearly seen in the high-frequency responses of the HRTFs at  $-90^\circ$ , which are substantially greater than in the HRTFs at  $-60^\circ$  at each source distance.

The effects of constructive and destructive interference on the contralateral hemisphere of the head are also seen in the ripples of the high-frequency HRTF response at  $-60^\circ$ . The

interference patterns produce a series of circularly symmetric peaks and nulls around the location of the bright spot, and the HRTFs at  $-60^\circ$  include four of these frequency-dependent nulls from 1 kHz to 15 kHz.

#### **6.1.5 Low-frequency diffraction**

The low-frequency response of the sphere-model HRTFs cannot be explained intuitively with the simple concepts described above. As noted before, the source-proximity effect only partially explains the low-frequency responses of the HRTF. The rest of the low-frequency responses are a result of diffraction by the head. Note that the diffraction effects tend to increase the low-frequency gain at the ipsilateral ear and increase the low-frequency attenuation at the contralateral ear. The magnitude of these diffraction effects increases as distance decreases.

#### **6.1.6 Low-pass filtering as a possible spectral distance cue**

The combination of the low-frequency diffraction effects in the ipsilateral hemisphere and the head-shadowing effects in the contralateral hemisphere produces a consistent relationship between the shape of the HRTF and source distance. At all source directions, the high-frequency response of the HRTF is lower relative to the low-frequency response of the HRTF when the source is close than when the source is more distant. The high-frequency response is generally 4-6 dB lower relative to the low-frequency response when the source is at 0.12 m than when the source is at 10 m (Figure 4). This relationship implies that a sound source at a fixed location relative to the head will appear to be low-pass filtered as the sound source approaches the head. Although this effect is modest, it could be used as a monaural distance cue in the near field, and it is consistent with previous observations that sound sources appear to “darken” in timbre as they approach the head.

#### **6.1.7 Contour plots of sphere-model HRTFs**

The left panels of Figure 5 provide a different perspective of the sphere-model HRTFs. Note that the more traditional HRTF plots shown in Figures 2 and 6 are essentially slices across frequency in these contour plots.

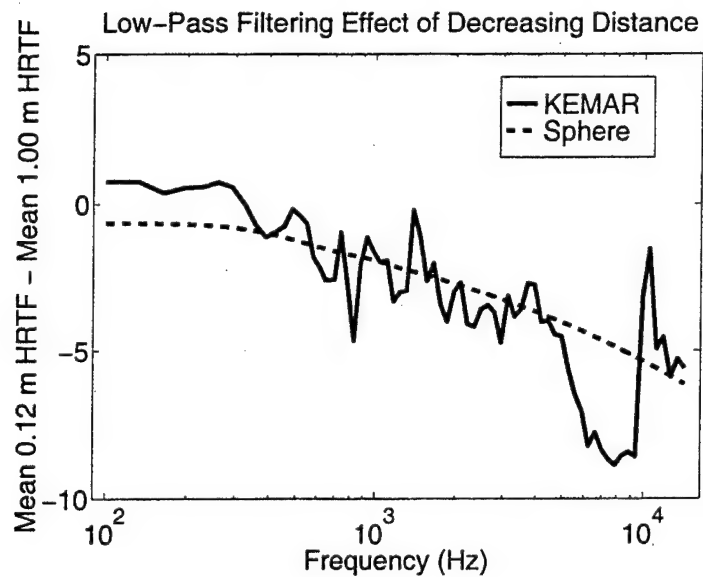


Figure 4: Low-pass filtering with distance. The sphere model and KEMAR HRTFs at 0.12 m and 1.0 m were averaged (in dB units) across all locations in the horizontal plane. The difference between the mean HRTFs at 0.12 m and 1.0 m illustrates that the monaural HRTF, on average, decreases in magnitude more quickly at high frequencies than at low frequencies as the source approaches the head. This effect is more pronounced in the KEMAR measurements than in the sphere model. The decrease in spectral content at high frequencies as distance decreases could potentially serve as a spectral distance cue in the near field.

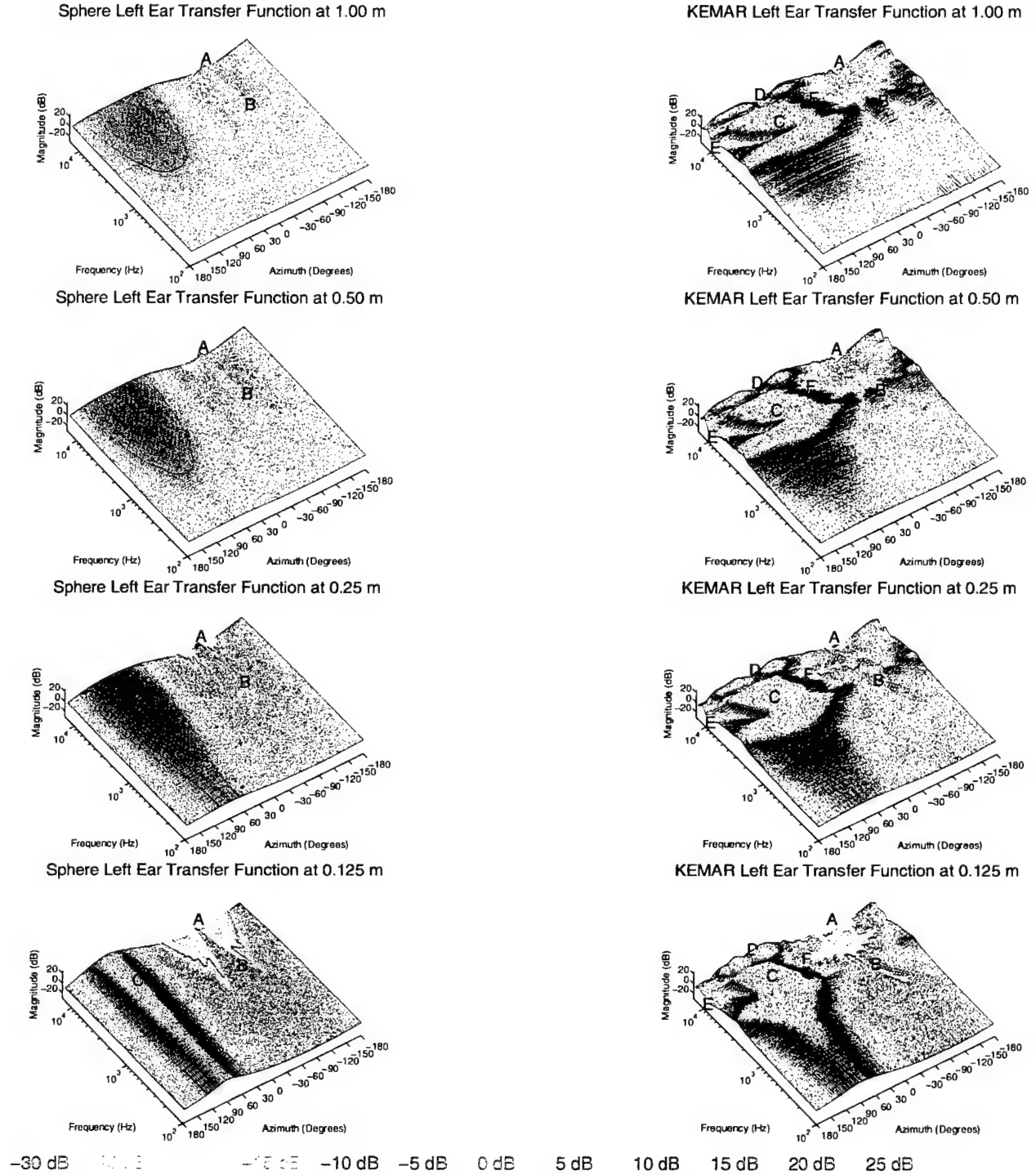


Figure 5: Surface contour plots of the monaural HRTFs predicted by the sphere model and measured with KEMAR. Azimuth is shown at  $3^\circ$  intervals, and frequency is shown at 100 Hz intervals from 100 Hz to 1 kHz, and at 1/12 octave intervals from 100 Hz to 15 kHz. The magnitude of the transfer function at each point is represented both by the height of the surface, shown on the Z-axis, and by the color, as shown by the legend across the bottom of the figure. In addition, contour lines are provided at 5 dB intervals ranging from -20 dB to 15 dB. Six reference points are present on the contour plots. Points A ( $-90^\circ$ , 15 kHz), B ( $-90^\circ$ , 2.5 kHz), and C ( $90^\circ$ , 2500 Hz) are shown in all the contour plots. Points D ( $75^\circ$ , 15 kHz), E ( $180^\circ$ , 6.5 kHz), and F ( $0^\circ$ , 7 kHz) are only on the KEMAR plots. Points B, C, and F are each located just above the surface of the plot.

These contour plots are particularly useful for viewing the periodic nature of the acoustic bright spot in the contralateral HRTFs. At 2500 Hz, there is a single peak around  $-90^\circ$  (location B). As frequency increases, this peak decreases in width and additional peaks form on either side, until at high frequencies the central peak is very sharp and is surrounded by multiple ridges on either side (location A). The increases in the high-frequency response of the ipsilateral HRTFs due to pressure doubling are also apparent in the contour plots (location C).

## 6.2 Measurements of Monaural Transfer Functions with the KEMAR Manikin

The monaural HRTFs measured with the KEMAR manikin are shown in Figure 6. Numerous studies examining HRTFs, both from manikins and from human listeners, are available in the literature, and the directional features of the far-field HRTFs are well documented. This discussion will focus on a comparison between the HRTFs calculated with the sphere model and measured with KEMAR, and on the distance-dependent features of the KEMAR HRTFs. Two important observations about the KEMAR HRTFs are detailed in the following sections:

- The overall shapes of the HRTFs are generally similar to those of the HRTFs calculated with the sphere model. At low frequencies (below 1 kHz), the sphere HRTFs are nearly identical to the KEMAR HRTFs. At higher frequencies, the KEMAR HRTFs diverge from the sphere model HRTFs, but the general direction and distance dependencies of the transfer functions are similar. The acoustic bright spot near  $-90^\circ$  shown in the sphere model is also apparent in the KEMAR transfer functions.
- The high-frequency features of the HRTFs are complex, particularly at the ipsilateral ear, but they appear to change systematically with source distance. In general, these features are compressed around the interaural axis as source distance decreases. This most likely results from the discrepancy between the location of the source relative to the ear and the location of the source relative to the center of the head.

### 6.2.1 Comparison of KEMAR measurements with sphere model

The sphere model best fits the KEMAR measurements at low frequencies (below 1 kHz). For comparison, the magnitude of the sphere model HRTF at 100 Hz is shown alongside the

## HRTFs for Left Ear Measured by KEMAR Manikin

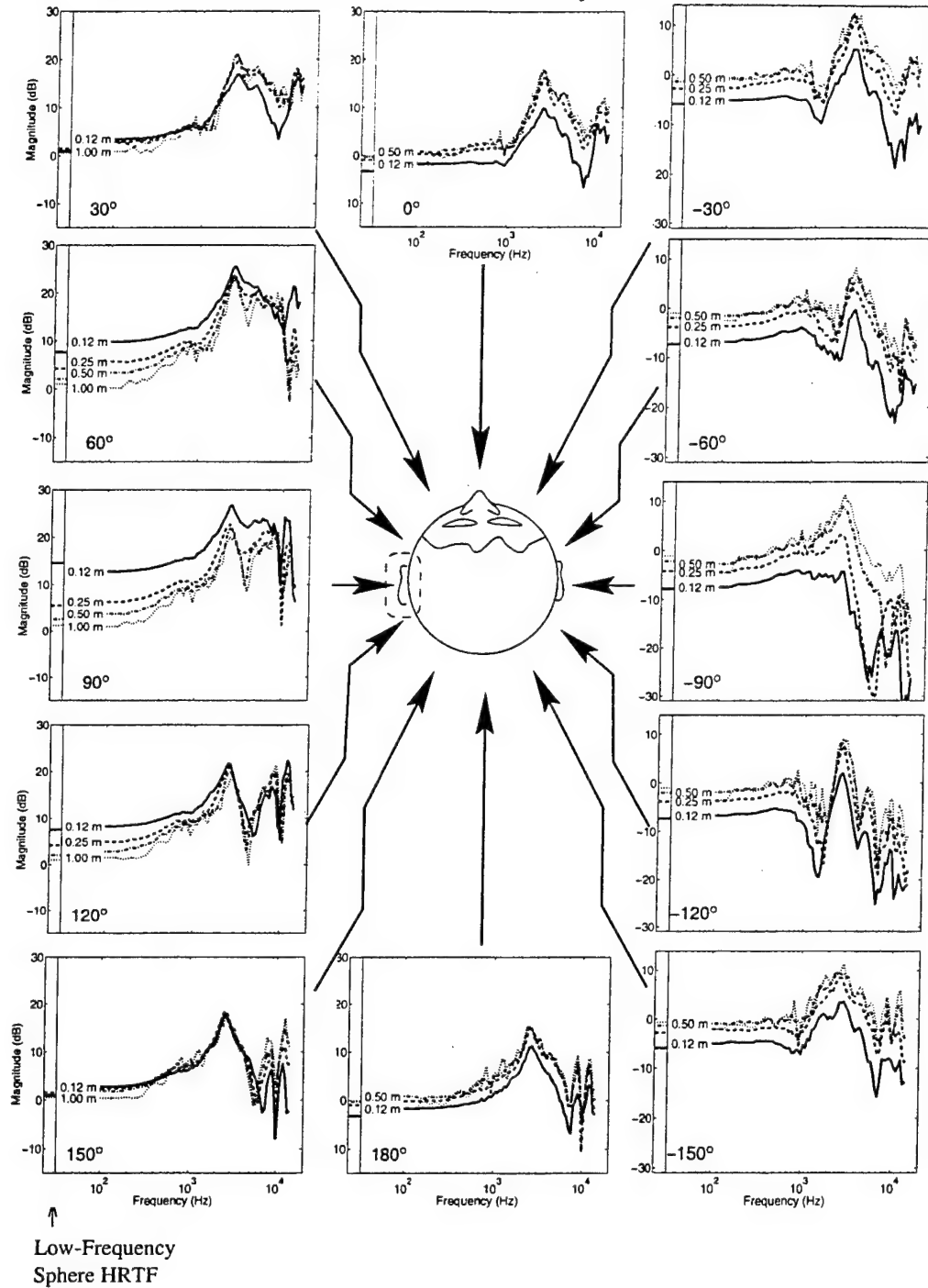


Figure 6: The monaural HRTF for horizontal-plane sources from 0.125 m to 10 m measured with the KEMAR manikin. The HRTFs were calculated by dividing the pressure at the left ear by the free-field pressure at the center of the head (see Equation 3). Results are provided at 30° intervals, and frequency is shown at 100 Hz intervals from 100 Hz to 1 kHz and at 1/12 octave intervals from 100 Hz to 15 kHz.

KEMAR measurements in Figure 6. The fit of the model to the measurements is best at the contralateral ear and worst near the boundary between the shadowed and unshadowed zones ( $30^\circ$  and  $-150^\circ$ ). There are also some discrepancies between the model and measurements at 0.12 m near  $90^\circ$ .

As frequency increases, the KEMAR HRTFs begin to diverge from the sphere model. At 2.9 kHz, the quarter-wavelength resonance of the ear canal causes a large peak in the KEMAR HRTFs at all directions and distances. At higher frequencies, the KEMAR transfer functions exhibit a complex series of direction-dependent peaks and notches which are derived from the geometry of the pinnae and are not reflected in the sphere model. Five major features of the sphere HRTFs are preserved in the KEMAR HRTFs.

- The magnitude of the HRTFs generally increases with frequency when there is a direct path from the source to the ear. In part, this feature is probably a result of the reflections that cause the 6 dB gain in the sphere model. The pinna, which is shaped like a cone, also provides some gain at high frequencies. Note that the overall high-frequency gain at the ipsilateral ear is greater in the KEMAR measurements than in the sphere model.
- The high-frequency responses of the HRTFs are attenuated when the ear is in the acoustic shadow of the head.
- The overall gain of the HRTFs increases as distance decreases when a direct path exists between the source and the ear, and the overall attenuation of the HRTFs increases as distance decreases when the ear is shadowed by the head. Note that, as in the sphere model, the ear is first shadowed by the head at  $30^\circ$  and  $150^\circ$  when the source is at 0.12 m, and that the ordering of the HRTFs at high frequencies reverses at these locations.
- Overall, the magnitude of the HRTF increases more rapidly at low frequencies than at high frequencies as the source approaches the head (Figure 4). Thus, the sound reaching the eardrums is effectively low-pass filtered as the source approaches the head. This effect is more dramatic in the KEMAR HRTFs than in the sphere model, and may serve as a monaural distance cue in the near field.
- Although its structure is more complicated, the acoustic bright spot seen in the sphere model is also found in the KEMAR measurements. This is best seen in the contour plots of the KEMAR measurements (Figure 5). The peak at intermediate frequencies occurs

slightly to the left of  $-90^\circ$  in the KEMAR measurements due to the asymmetries of the manikin head (location B). At higher frequencies, the periodic interference pattern around the bright spot is seen in the KEMAR measurements, but it is more erratic than in the sphere model. Note that one notch from the interference pattern appears to extend into the ipsilateral hemisphere at 6.5 kHz (locations E and F).

### **6.2.2 Distance dependent changes in the high-frequency ipsilateral HRTF**

The high-frequency features of the KEMAR transfer function are complex and are related to the geometric features of the manikin head and torso and, particularly at high frequencies, the folds and cavities of the external ear. Some of these features are present in the HRTFs of a wide variety of humans and manikins, and change consistently with the direction of the source. Shaw (1974) provides an excellent analysis of far-field HRTFs that describes several such features. In this analysis, two features of the high-frequency HRTF will be analyzed as a function of distance to provide insight into the behavior of the HRTF in the near field.

Both features are visible in the contour plots of Figure 5. The first is a sharp notch located just to the left of location C. This notch extends from  $180^\circ$  to  $60^\circ$  in the 1.0 m measurements, and decreases in frequency from 7 kHz to 4 kHz as azimuth increases. As the source distance decreases, note that this notch becomes shorter and decreases in frequency with azimuth more rapidly. This notch is visible at approximately 7 kHz at  $180^\circ$  in Figure 6 and decreases in depth and in frequency as azimuth moves to  $60^\circ$ .

The second feature is a pair of peaks in the HRTFs at 14 kHz, located at  $90^\circ$  and  $125^\circ$  at 1 m. These peaks are bordered by a deep notch across all ipsilateral locations at 9 kHz, and are separated by a deep notch at approximately  $60^\circ$  (location D). As distance decreases, these peaks are compressed around  $90^\circ$ .

A better view of these features is provided by Figure 7. In this figure, the notch is marked by a white arrow, and the larger of the two peaks is encircled by a dotted white line (the smaller peak is just below the dotted box). From this figure, note again that the notch decreases dramatically in length as distance decreases from 1.0 m to 0.12 m. In addition, the rate at which the frequency of the notch increases with angle is greatest at 0.12 m. Note in Figure 6 that, at  $120^\circ$  and  $150^\circ$ , the notch is consistently at a higher frequency at 0.12 m than at the other measured distances. In effect, the notch has been pushed away from  $90^\circ$  at the closest distance.

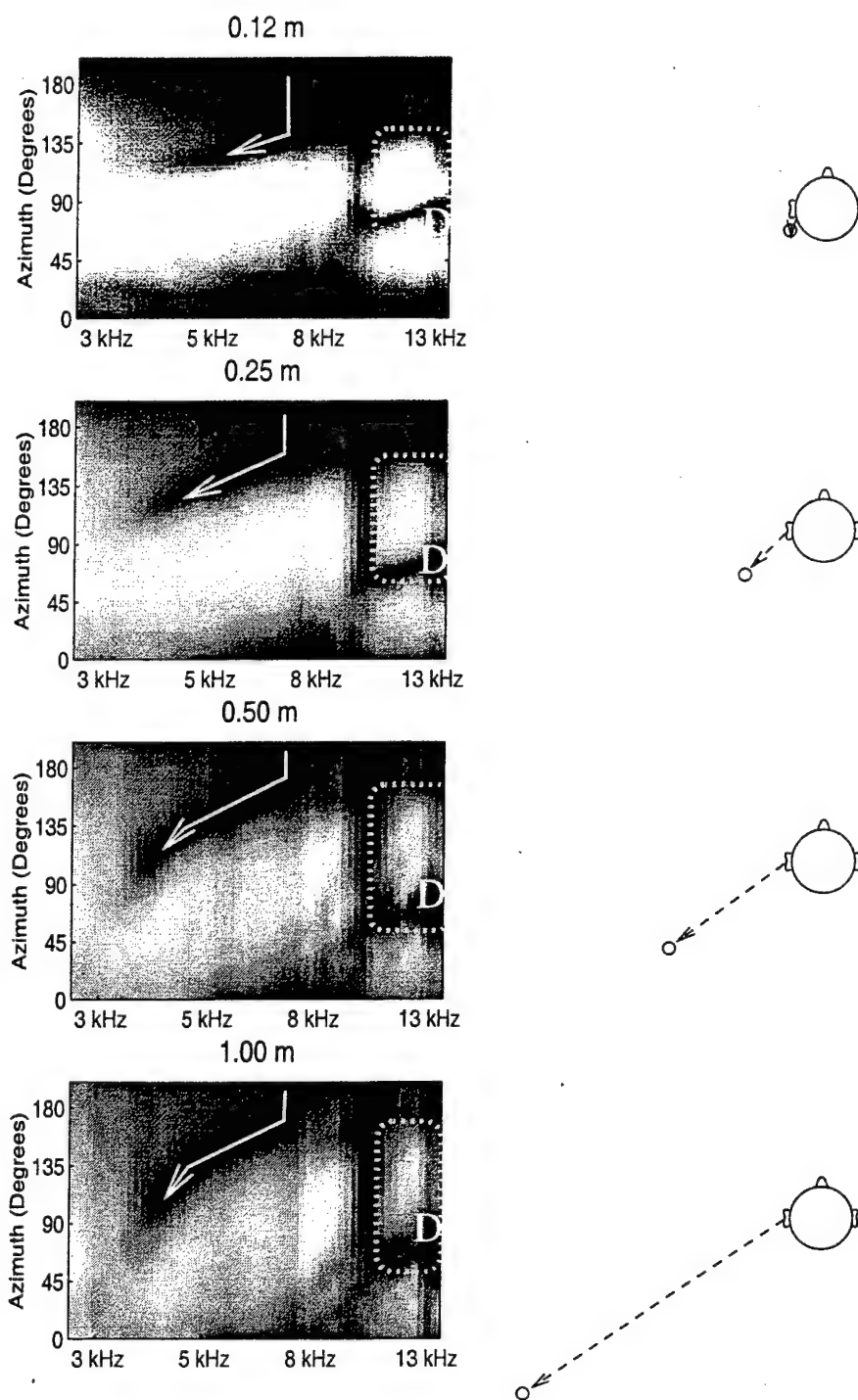


Figure 7: Detailed view of ipsilateral HRTF features at high frequencies. In these plots, darker features indicate a lower magnitude in the HRTF. The notch described in the text is shown by the white arrows, and the larger peak discussed in the text is surrounded by the dotted line. The right panels show the orientation relative to the ear of a source located at 120° relative to the head at each distance.

There is no obvious explanation for the behavior of this notch, but it most likely results from a reflection from the head or torso destructively interfering with the direct signal at the location of the ear. The notch would be pushed away from  $90^\circ$  because, as the ear rotates toward the source, the ratio of the length of the direct path to the length of the reflected path decreases and the direct signal becomes stronger than the reflected signal and, consequently, less susceptible to destructive interference. The changes in frequency result from changes in the path length of the reflection due to the geometrical configuration of the head and a nearby source.

The second feature, the pair of peaks in the HRTF at 13 kHz, also changes systematically with distance. At 1.0 m, the peaks are relatively broad, extending from  $0^\circ$  to nearly  $180^\circ$  in azimuth. As distance decreases, the peaks progressively narrow and, at 0.12 m, they extend only from  $20^\circ$  to  $140^\circ$ . The sharpness of the peaks also increases as distance decreases. The deep notch separating the peaks moves from approximately  $60^\circ$  at 1.0 m to nearly  $80^\circ$  at 0.12 m. It appears that these peak features are compressed around  $90^\circ$  as distance decreases.

The distance dependencies of the high-frequency peaks are easily explained geometrically by the discrepancy between the angle of the source relative to the ear and the angle of the source relative to the head. At high frequencies, the shape of the ear is the primary determinant of the features of the HRTF, and the response of the ear is governed by the direction of the sound waves impinging on the pinna. The right side of Figure 7 shows how a source located at  $120^\circ$  relative to the head changes in orientation relative to the ear as distance decreases from 1.0 m to 0.12 m. At the furthest distance, the direction relative to the ear is approximately equal to the direction relative to the head; but, as distance decreases, the angle relative to the ear increases substantially and, at 0.12 m, the source at  $120^\circ$  is located at  $180^\circ$  relative to the ear. As a result, the high-frequency features based on orientation relative to the ear are compressed around the interaural axis.

## 7.0 INTERAURAL INTENSITY DIFFERENCES

Interaural intensity differences were calculated at 0.125 m, 0.25 m, 0.50 m, and 1.0 m with the KEMAR HRTFs, and at each of these distances plus 10.0 m with the sphere model HRTFs. The data are shown in polar form in Figure 8. The important characteristics of the IIDs can be summarized as follows:

- The IID is always 0 dB in the median plane and generally increases as the source moves lateral to the head. This result follows directly from the directional dependence of the monaural HRTFs, which increase in magnitude as the ear rotates toward the source and decrease in magnitude as the ear rotates away from the source.
- The IID generally increases as frequency increases. This behavior results from the tendency of the monaural HRTF to increase with frequency when there is a direct path from the source to the ear and to decrease with frequency when the ear is shadowed by the head. Both effects contribute to the enlarged IID at high frequencies.
- The IID increases as distance decreases, and increases dramatically as the source distance drops below 0.5 m. This distance dependence occurs because the magnitude of the monaural HRTF increases as distance decreases at the ipsilateral ear, and decreases as distance decreases at the contralateral ear.
- The acoustic bright spot directly opposite the location of the source causes a local minimum in the IID near  $\pm 90^\circ$  at intermediate frequencies (1500 Hz and 3000 Hz). In the sphere plots, this minimum also occurs at higher frequencies, but in the KEMAR HRTFs the irregular shape of the head causes the bright spot to break into an erratic series of peaks and nulls which influence the IID around  $90^\circ$ . At 500 Hz, the bright spot does not significantly influence the IID.

As with the monaural HRTFs, the sphere model most accurately reflects the behavior of the KEMAR measurements at low frequencies. The IID at 500 Hz is a very smooth function both for the sphere model and the KEMAR measurements. At higher frequencies, the asymmetries of the KEMAR head cause the KEMAR IID to deviate from the sphere model

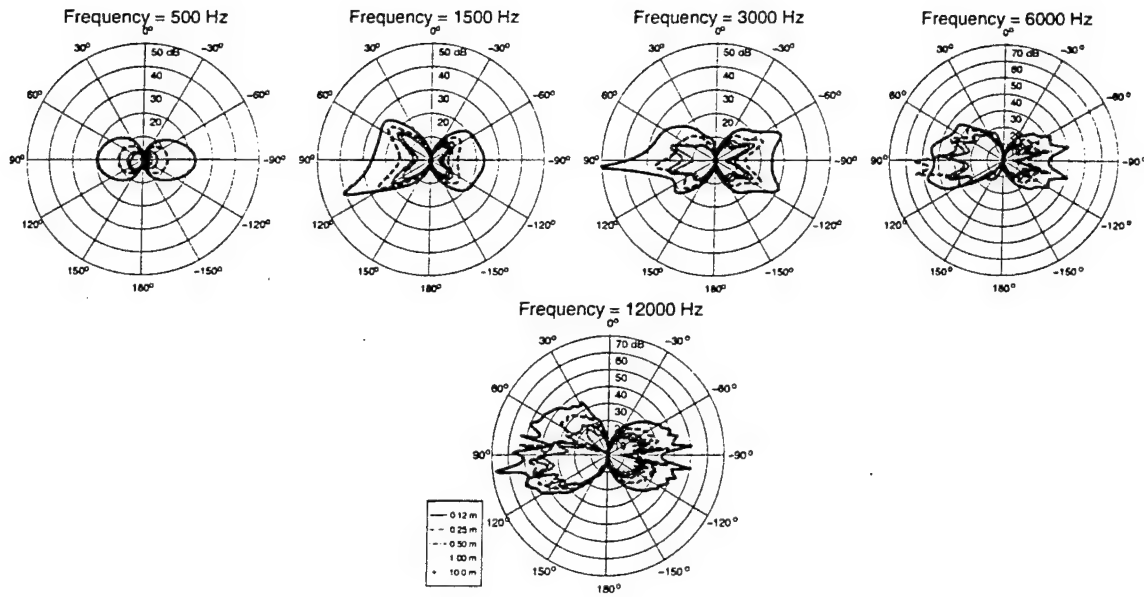


Figure 8: Interaural intensity differences. In these polar plots, the location of the source in azimuth is represented by angle, and the magnitude of the IID (in dB) is represented by the radius at each angle. Results are provided at five frequencies, ranging from 500 Hz to 12 kHz. The left side of each plot shows the IID from the KEMAR measurements, while the right side shows the IID calculated from the sphere model. Note that the scale in the plots at 6000 Hz and 12000 Hz is larger than in the lower-frequency plots.

IID. At these frequencies, the KEMAR IIDs tend to be significantly larger than the IIDs predicted by the sphere model. This discrepancy is primarily the result of the directional properties of the pinna, which provides a significant amount of mid- to high-frequency gain at the ipsilateral ear.

## 8.0 INTERAURAL TIME DELAYS

The interaural time delay for a source in the horizontal plane was calculated at 0.125 m, 0.25 m, 0.50 m, 1 m, and 10 m with the sphere model and measured at 0.12 m, 0.25 m, 0.50 m, and 1 m with the KEMAR manikin (Figure 9). In each case, the time delay was calculated from the unwrapped phase of the difference spectrum between the left and right ears. The time delay was determined from the slope of the line best fitting the unwrapped phase of the difference spectrum from 100 Hz to 6.5 kHz. Positive time delays indicate a phase lag at the right ear, and negative time delays indicate a phase lag at the left ear.

The time delays from the sphere model are necessarily symmetric across both the median and frontal planes. At 10 m, the time delay peaks at approximately  $700\ \mu\text{s}$  at  $90^\circ$ . As the distance decreases, the magnitude of the time delay increases slightly. This increase is most dramatic at  $90^\circ$  and  $-90^\circ$ , where the time delay increases by about  $100\ \mu\text{s}$  as the distance decreases from 10 m to 0.125 m. The majority of this increase occurs as the source moves from 0.25 m to 0.125 m, and the remainder from 0.50 m to 0.25 m. There is virtually no dependence between the time delay and distance beyond 0.50 m. When the source is not near the interaural axis, the time delays do not vary significantly with distance.

The KEMAR time delay measurements are similar, except that the asymmetrical shape of KEMAR's head is readily apparent. This asymmetry causes the time delay to drop off more rapidly in the rear hemisphere than in the front hemisphere. Also, the time delays with the KEMAR manikin exhibit a much broader peak when the source is near the interaural axis. As with the time delays predicted by the sphere model, the KEMAR time delays increase by approximately  $100\ \mu\text{s}$  as distance decreases from 1 m to 0.12 m. One slight difference between the KEMAR time delays and those predicted by the sphere model is the mild increase in the magnitude of the time delay in the front hemisphere at 0.12 m. The sphere model predicts almost no increase in the delay except in the immediate vicinity of the interaural axis.

Both the sphere model and the KEMAR measurements indicate that the dependence of the interaural time delay on distance is substantially weaker than the dependence of the interaural intensity difference on distance. At the closest distances, the magnitude of

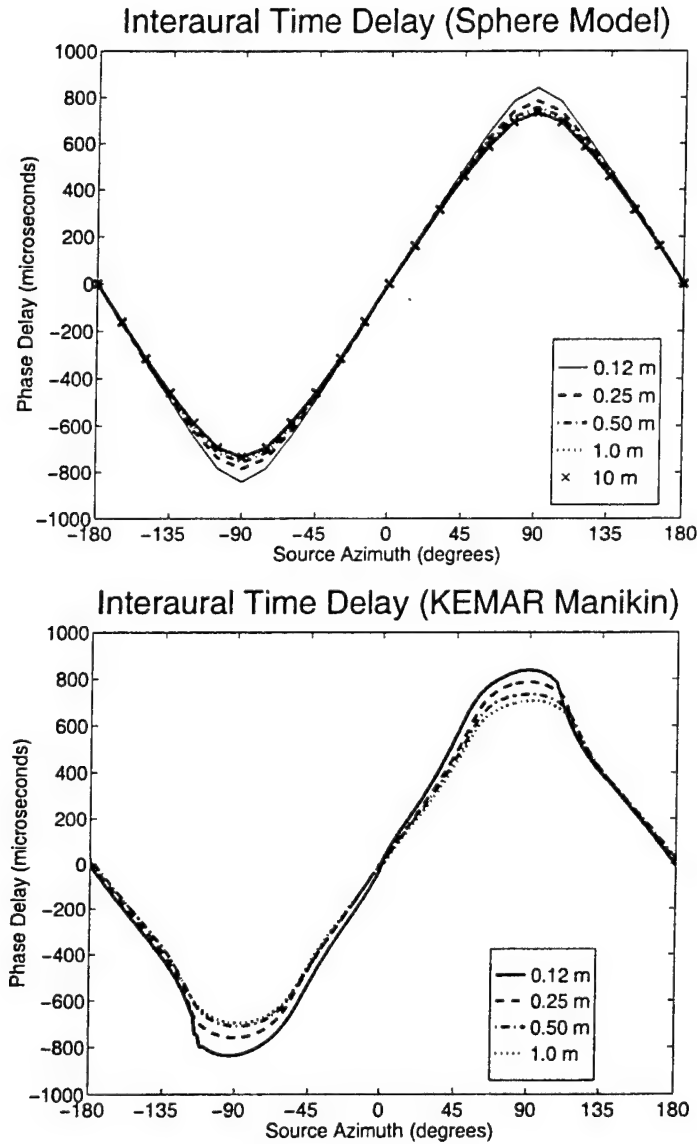


Figure 9: Interaural time delays. The delay was determined from the best linear fit of the unwrapped phase difference between the left and right ears (see text for details). Positive delays indicate a lag at the right ear, and negative delays indicate a lag at the left ear.

the IID increases dramatically, while the ITD never increases by more than 10-12%. The reason for this discrepancy is simple: the time delay depends on the arithmetic difference between the distance from the source to the ipsilateral ear and the distance from the source to the contralateral ear. Ignoring the effects of diffraction by the head, the IID depends on the ratio of the distance to the ipsilateral ear to the distance to the contralateral ear. As the source approaches the head, the ratio of distances to the ipsi- and contralateral ears increases much faster than the absolute difference between the distances, so the IID increases more dramatically than the ITD. As discussed later, this disparity between the distance dependence of IID and ITD is even greater if perceptual considerations are considered.

## 9.0 ELEVATION EFFECTS ON NEAR-FIELD HRTFS

The high-frequency features of the monaural HRTF at the ipsilateral ear change substantially with elevation. This is illustrated in Figure 10, which shows the high-frequency features of the monaural HRTF at three elevation locations. The pattern of peaks and notches in the transfer functions is significantly different at each elevation location. At  $+30^\circ$  elevation, for example, there is a notch at approximately 7 kHz in the HRTF that stretches across the entire ipsilateral hemisphere which is not found at either of the other two elevations. At  $+30^\circ$  elevation, there is a wide null near  $0^\circ$  azimuth at 8 kHz. Similar patterns have been reported in previous HRTF studies (Carlile & Pralong, 1994; Shaw, 1974) and they will not be discussed in detail here. It is apparent, however, that these patterns do vary significantly with elevation, and that they could provide a salient cue for evaluating the elevation of a sound source.

These patterns are apparently relatively independent of source distance. Careful observation reveals that the features of the HRTFs are considerably more consistent across the rows of Figure 10, which represent different elevation values, than across the columns, which represent different distances. The general pattern of features at each elevation is clearly recognizable at all three measured distances. If this result is generally true at all elevations, it would imply that elevation cues are roughly independent of distance and that the same mechanisms that allow elevation perception in the far field may also be used in the near field.

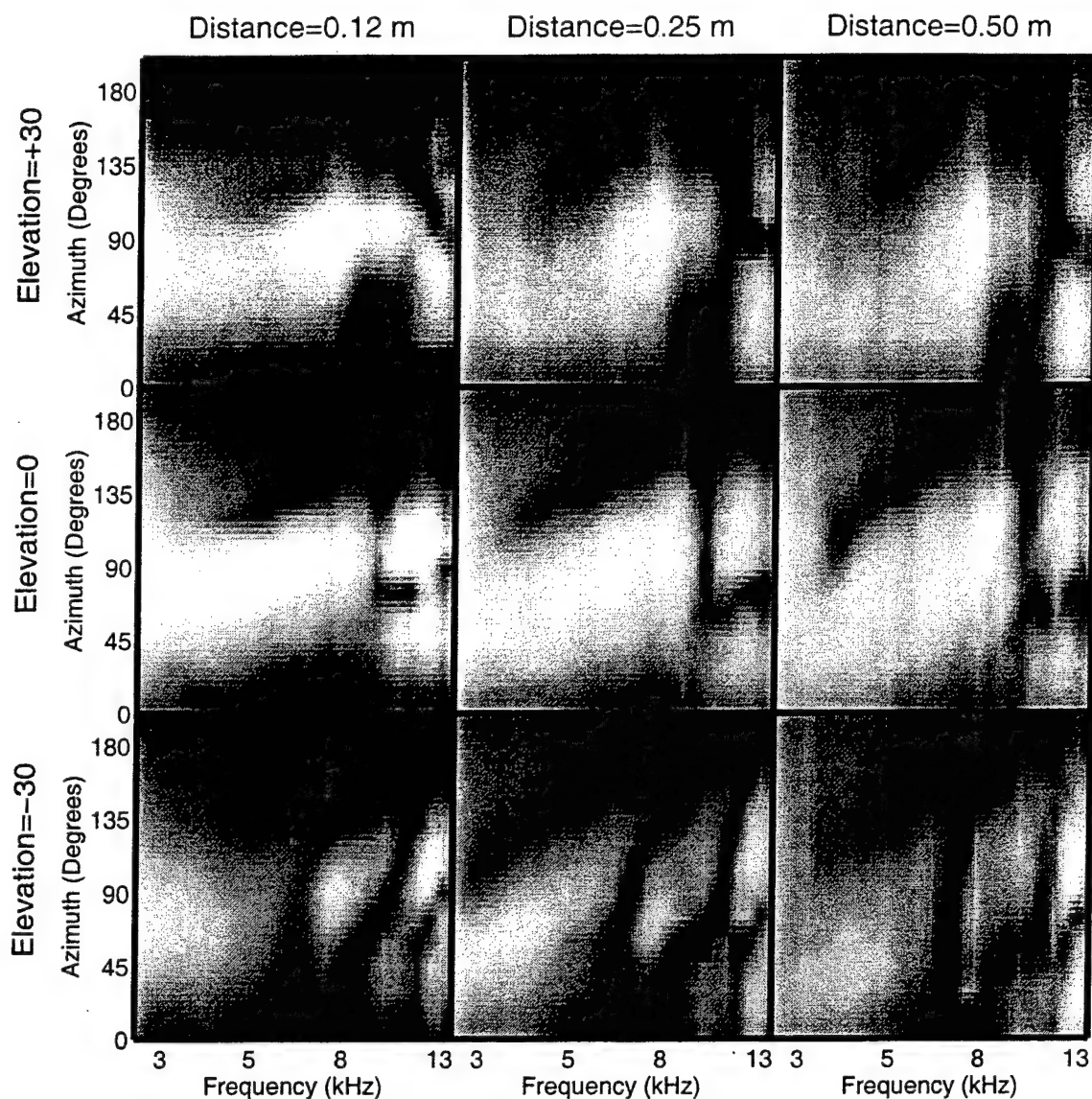


Figure 10: Surface plots of the left-ear monaural HRTFs measured with the KEMAR manikin at +30°, 0°, and -30° in elevation. As in Figure 7, the darkness of the plots indicates the magnitude of the HRTF at each point: brighter areas indicate greater gain in the HRTF. The plots are limited to high frequencies at the ipsilateral ear, where the most salient elevation cues occur, and are shown with 1/12th octave resolution in frequency and 15° resolution in azimuth (interpolation has been used to smooth the plots).

## 10.0 PINNAE EFFECTS FOR VERY NEAR SOURCES

The geometry of the pinna is complex, and it is difficult to predict how its response will change when a source is close to the ear. In order to isolate the contribution of the pinna to the near-field HRTF, the transfer function at the right ear was measured for sources located just outside the ear (at  $-90^\circ$  azimuth) with and without the pinnae. Measurements were made at distances of 2.5 cm, 3.75 cm, 5 cm, 7.5 cm, and 10 cm, measured from the opening of the ear canal rather than the center of the head. At each location, an initial measurement was made with the standard KEMAR pinna; then a second measurement was made with the removable pinna of the KEMAR manikin replaced by a flat rubber sheet flush with the surface of the head and the opening of the ear canal.

The HRTF at each position was calculated by dividing the frequency response measured at the ear by the free-field pressure at the center of the head if the manikin were removed. Calibration measurements were unavailable at these distances, so the free-field signal was calculated by scaling the free-field transfer function measured at 12 cm in accordance with the inverse relation between pressure and distance.

The features of the monaural HRTFs with the pinna attached (upper panel of Figure 11) are similar to those measured for the left ear at  $90^\circ$ . The gain of the transfer function generally increases as distance decreases, but the increase is considerably larger at low frequencies than at high frequencies. Note the large discontinuity between the HRTFs at 3.75 cm and 2.5 cm at high frequencies. At this point, the source is surrounded by the concha and the HRTF changes dramatically.

When the pinna is removed, the variations in the HRTF with distance are more systematic (middle panel of Figure 11). Notice that the increase in gain as the source approaches the head decreases as frequency increases from 100 Hz to 2 kHz, but is roughly independent of frequency beyond that point. This pattern is similar to the HRTF predicted by the sphere model for a source at  $90^\circ$  (Figure 2). Note that the elimination of the pinna effectively decreases the length of the ear canal by about 20%, resulting in a 500 Hz increase in the quarter-wavelength ear-canal resonance when the pinna is removed (label ECR on the graphs).

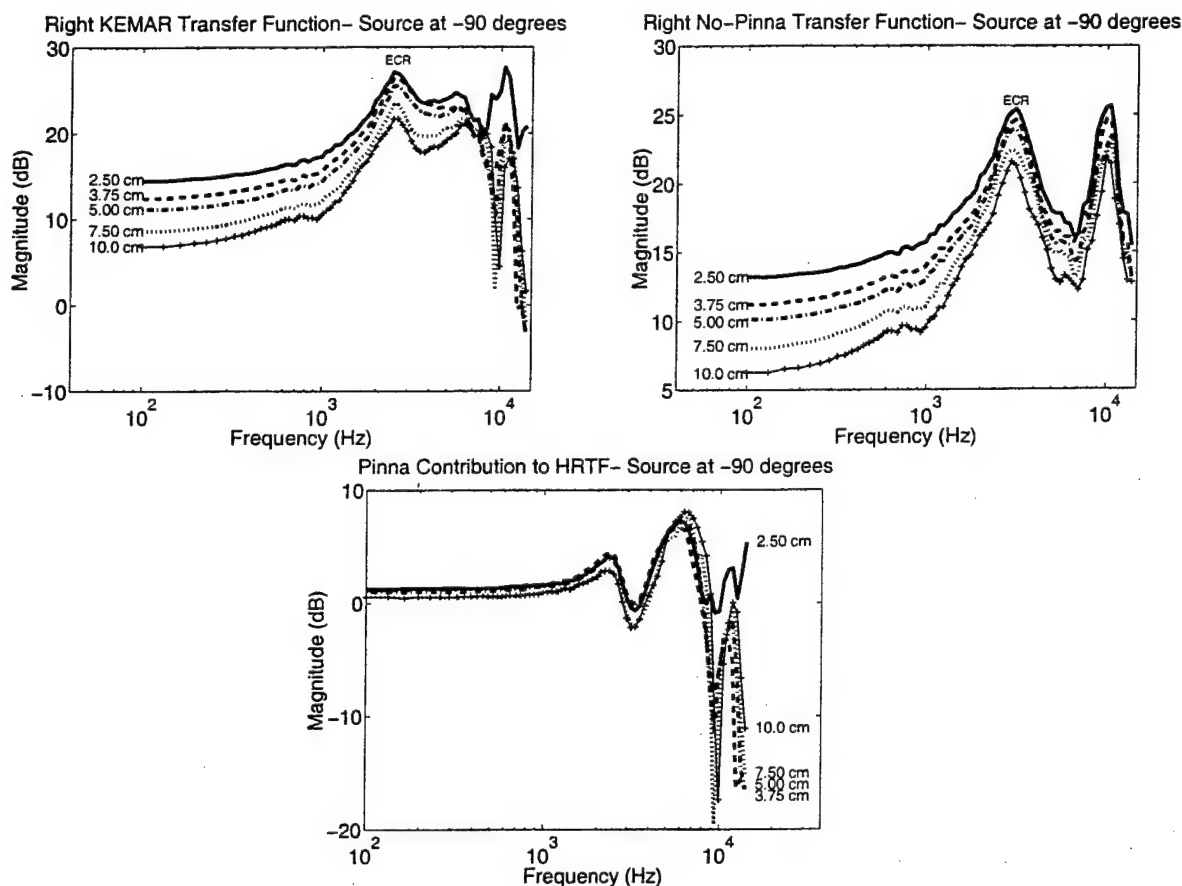


Figure 11: The contribution of the pinna to the HRTF for a nearby source at  $-90^\circ$ . Measurements were made with the right ear of the KEMAR manikin both with the standard pinna attached (shown in the first panel) and with the pinna replaced by a flat rubber sheet with an opening at the ear canal (shown in the second panel). The third panel shows the ratio of the transfer function with the pinna to the transfer function without the pinna. All distances were measured from the surface of the head, rather than the center of the head. The label ECR represents the location of the ear canal resonance.

The bottom panel of Figure 11 shows the ratio of the HRTFs with and without the pinna, which represents the contribution of the pinna to the HRTF. The pinna contribution is roughly independent of distance at frequencies up to 8 kHz. The peak and notch near 3 kHz reflect the 500 Hz increase in the frequency of the ear-canal resonance when the pinna is removed. At higher frequencies, the contribution of the pinna changes only modestly with distance, with the exception of the large increase in the high-frequency response of the pinna as the source moves from 3.75 cm to 2.5 cm.

These results contrast with those of Shaw and Teranishi (1968). They measured the frequency response of an outer-ear replica mounted in a rigid plate for a nearby point source, and found that the low-frequency gain provided by the outer ear increased with decreasing distance and that the high-frequency gain of the ear decreased with decreasing distance. The low-frequency effect described by Shaw is not seen in our data, and the high-frequency effect is found only at distances greater than 2.5 cm.

Although data are available only for sources at 90° azimuth and 0° elevation, both Shaw and Teranishi's measurements and our measurements indicate that the response of the ear is roughly independent of distance when the source is located more than 4 cm from the ear. Thus it appears that most of the distance-dependent changes in the near-field HRTFs result from sound diffraction by the head and torso and the geometric orientation of the source relative to the ear, and not from changes in the acoustic behavior of the pinna in the near field.

## 11.0 COMPARISON OF SPHERE MODEL AND KEMAR MEASUREMENTS

The accuracy of the sphere model at low frequencies is displayed in Figure 12, which shows the distance dependencies of the IID and ITD from the sphere model and from the KEMAR measurements. In general, the measured data fit the predictions of the sphere model well. The only exception is at  $90^\circ$  in the 3 kHz plot, where the IID is significantly larger in the KEMAR measurements. Two factors contribute to this result. First, the resonance of the ear canal is greater on the ipsilateral side than on the contralateral side, resulting in a net increase in the IID with the manikin. Second, the acoustic bright spot, which causes an increase in pressure at the contralateral ear and decreases the IID, is less pronounced with the irregularly shaped head of the manikin than with a perfectly spherical head. Although the model is considerably less able to predict the actual HRTF at higher frequencies, it is a valuable tool for predicting many of the features of the near-field HRTF.

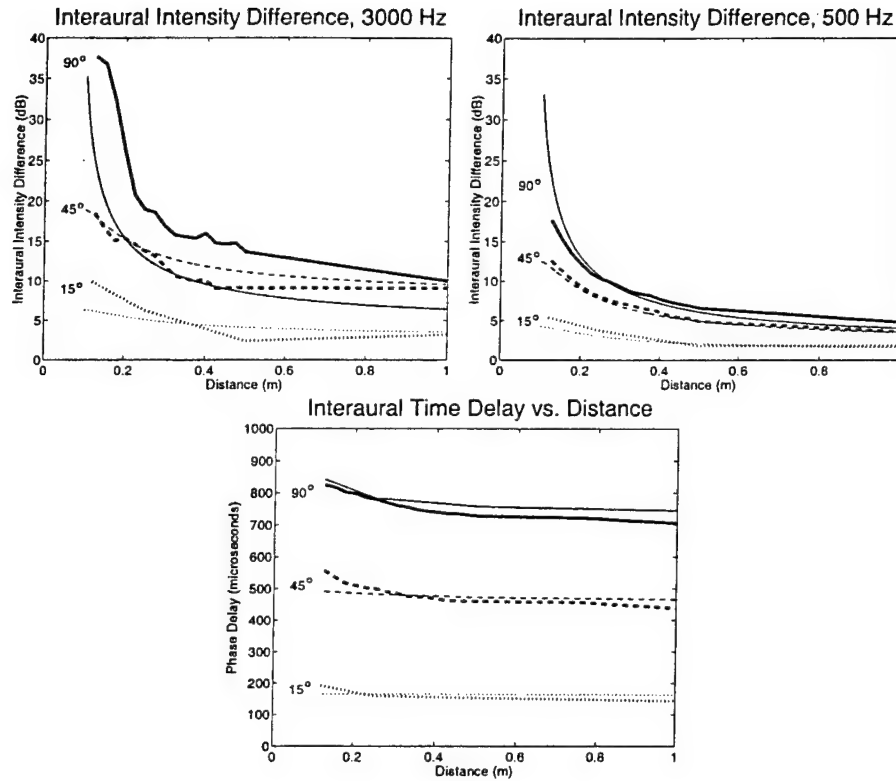


Figure 12: Comparison of sphere model and KEMAR measurements. The bold lines are the KEMAR measurements, and the thin lines are the corresponding predictions by the sphere model. The KEMAR measurements at 45° and 90° are shown every 2.5 cm from 0.12 cm to 0.50 m, and at 1 m; the other data are shown only at 0.12 m, 0.25 m, 0.50 m and 1 m.

## 12.0 PERCEPTUAL IMPLICATIONS OF THE NEAR-FIELD HRTFS

Figure 12 is also useful for analyzing the perceptual implications of the distance-dependent attributes of the HRTF in the near field. The IIDs increase rapidly as distance decreases, especially at distances less than 0.5 m, while the ITDs increase only slightly at distances less than 1 m. The disparity between the distance dependence of the ITD and IID in the near field is even larger when perceptual issues are considered. Hershkowitz and Durlach (1969) found that listeners could discriminate changes in IID on the order of 0.8 dB, so the changes in IID from 0.125 m to 1 m span a range of up to 15 JNDs at 500 Hz and 30 or more JNDs at 3 kHz. The JND for ITD was approximately 15  $\mu$ s at ITDs below 400  $\mu$ s and increased rapidly for ITDs greater than 400  $\mu$ s, so the changes in ITD in the near-field span, at most, a few JNDs. Therefore subjects can be expected to perceive large changes in IID as the distance of a nearby source changes while the perceived ITD remains relatively constant.

The combination of perceptually invariant ITDs and strongly distance-dependent IIDs suggests a possible strategy for determining the distance of a nearby source in the horizontal plane based on binaural information from the HRTFs. The ITD information, which is relatively independent of distance, could be used to identify the azimuthal direction of the sound source. Once the source direction in azimuth is known, the systematic dependence of IID on distance could be used to estimate the distance of a sound source, provided the source is outside the median plane (where the IID is near zero at all distances). This model of near-field binaural distance perception would predict the following characteristics in near-field distance estimation performance:

1. Distance accuracy would be greatest for sources on the left or right side of the listener and worst for sources directly in front or behind. Since the variation in IID with distance is greater for lateral sources than for sources in the median plane, listeners would have less resolution in distance perception near 0° and 180°.
2. The percentage JND in distance at a fixed azimuth would increase as distance increases. In Figure 12, the slope of the curve relating IID to distance increases substantially as

the source approaches the head. Thus the percent decrease in distance necessary to produce a fixed increase in IID decreases as distance decreases. If the JND in distance is defined as the percent decrease in distance necessary to produce a single JND in interaural intensity, the distance JND will decrease with distance.

3. Provided the source is sufficiently broad-band to allow the listener to perceive both ITD and IID information, distance perception would not depend on the spectral shape or intensity of the source. The key advantage to binaural depth perception, in contrast to depth perception based on intensity, spectral cues, or reverberation, is that the binaural information is derived from the difference signal between the left and right ears and does not depend on the characteristics of the source (except its frequency range).

Note that this model is similar in concept to one suggested by Hirsch (1968), further explored by Greene (1968) and expanded by Molino (1970). This model demonstrated the possibility of determining the distance of a sound source based on the relationship between the IID and ITD. Hirsch's model, which ignored diffraction by the head and assumed the ears were detectors in free space, predicted that distance could be calculated directly from the ratio of ITD to IID. Molino's expanded model, based on a spherical head, required that the azimuth location of the source to be known *a priori*. Greene and Molino used threshold data for ITD and IID to calculate the predicted accuracy of distance perception using this model and, predictably, found that distance perception in the far field would be very poor. Molino noted that predicted accuracy would be much greater in the near field, due to the dramatic increase in IIDs in that region. The present data indicate not only that the changes in IID in the near field are easily perceptible, but also that the relative invariance of the ITD in the near field may allow listeners to determine azimuth directly from the ITD without external knowledge about the direction of the source.

The situation becomes more complex outside the horizontal plane, but it is still possible to determine the azimuth, elevation, and distance of a sound source from the HRTF. Recall that the KEMAR measurements at  $30^\circ$  and  $-30^\circ$  indicate that the high-frequency, elevation-dependent features in the HRTF, which are believed to be important in localizing elevation, are roughly independent of distance. If these high-frequency cues could be used to determine the elevation of the source, and compensate for the elevation-dependent changes in the IID and ITD, the azimuth and distance of the source could still be accurately determined from interaural cues. This model would imply greater distance accuracy in the horizontal plane than the median plane, as IIDs decrease as a source moves directly above or below a listener.

Another possible strategy for determining the distance of a nearby source involves the disparity between the orientation of the source relative to the head and the orientation of the source relative to the ear. The ITDs, as well as the low- to mid-frequency IIDs, are determined primarily by the orientation of the source relative to the center of the head. However, the high-frequency features of the HRTF are largely a result of the geometric properties of the pinnae, and therefore are governed by the orientation of the source relative to the ear. This causes a compression in the spatial locations of high-frequency HRTF features in the near field (Figure 7). If listeners were able to determine the azimuth of a sound source relative to the ear with high-frequency spectral cues, and relative to the head with low-frequency IIDs, they conceivably could use the two values to triangulate the distance of the source.

There is some evidence that listeners can use pinna-based cues to determine the direction of a sound source. Studies examining monaural localization ability have shown that listeners have some ability to identify the location of a broadband sound when one ear is occluded by an ear-plug and muff (Wightman & Kistler, 1989; Butler, Humanski, & Musicant, 1990; Oldfield & Parker, 1986) or congenitally impaired (Slattery & Middlebrooks, 1994). The apparent position in azimuth of monaural narrow-band stimuli is related to the direction-dependent gain of the pinna at the center frequency of the signal (Rogers & Butler, 1992; Butler, 1987; Musicant & Butler, 1984; Butler et al., 1990; Belendiuk & Butler, 1977). Azimuth localization based on pinna cues does not, however, appear sufficiently accurate to allow accurate perception of distance via triangulation. Therefore, it is unlikely that subjects are able make distance judgments based on the geometric location of the source relative to the head and ear.

A final possible distance cue indicated by the HRTF measurements is a slight increase in the relative low-frequency gain as distance decreases. As distance decreases, the gain of the HRTF increases more at low frequencies than at high frequencies at the ipsilateral ear, and the attenuation due to head shadowing increases more at high frequencies than at low frequencies at the contralateral ear. As a result, the signal reaching the eardrums from a nearby source is effectively low-pass filtered as the source approaches the head (Figure 4). This low-pass filtering may explain the "darkening" of a very near sound source reported by von Békésy (1960). Von Békésy observed that the particle velocity of a spherically-radiating sound wave is increased relative to the pressure of the wave at distances less than a fraction of a wavelength from the source, and suggested that this might produce an increase in low-frequency energy for a nearby source (since the low-frequency components of the sound are

fewer wavelengths distant from the source than higher frequency components) (Coleman, 1963). Von Békésy's subjects reported that integrated (effectively low-pass filtered) sound bursts were perceived closer than differentiated (high-pass filtered) stimuli. There is no evidence, however, that the ear can directly perceive the velocity of a sound source, so von Békésy's explanation of this effect is suspect. Begault (1987) also noted the "darkening" in the timbre of very near sound sources, and suggested that the tendency to perceive larger increases in loudness at low frequencies than at higher frequencies from an equivalent increase in sound pressure level (the so-called Fletcher-Munson curve) might provide an explanation. Since the pressure level increases at all frequencies as the source approaches the ear, the Fletcher-Munson curve suggests that the perceived increase in loudness would be greater at low frequencies. The "darkening" of near-field stimuli reported in the literature is probably a combination of the boost in low-frequency gain due to near-field acoustic effects shown in the HRTFs and the non-uniform perception of increasing loudness across frequencies.

The implications of the near-field properties of the HRTF in directional localization are less obvious than those in distance localization. There is evidence that low-frequency time delays (Wightman & Kistler, 1992) dominate the perception of azimuth when they are available. Since time delays vary only slightly as distance decreases, it is not likely that azimuth perception will be significantly different in the near field than in the far field when the spectrum of the source extends into low frequencies. If time delay information is not available from low-frequency time delays or high-frequency envelope delays (Middlebrooks et al., 1989), it is possible that localization ability may degrade substantially. Without time delay information, there is no obvious mechanism for determining the relative contributions of distance and direction to the IID. In other words, there is no way to determine whether a certain IID is the result of a distant sound source near the interaural axis or a nearby sound source near the median plane. The consequences of this confusion on localization performance are unclear. The increase in IIDs at close distances will, however, decrease the JND in azimuth to the extent that the JND is limited by the change in IID. It is also possible that the change in IID with head orientation could provide a strong dynamic distance cue when exploratory head motions are allowed. In addition, the spatial remapping that occurs in the high-frequency pinnae cues in the near field may introduce a lateral bias if the auditory system is using these cues in the perception of azimuth.

The high-frequency pinna cues that are believed to determine elevation were found to be roughly independent of distance over the limited range of elevations measured. These cues are believed to dominate the perception of elevation, so it is unlikely that the localization of elevation is strongly dependent on distance in the near field.

## 13.0 CONCLUSIONS

The major conclusions of this study can be summarized as follows:

1. The dominant distant-dependent feature of both the sphere model and KEMAR HRTFs in the near field is an increase in the interaural intensity difference with decreasing source distance across all frequencies. When a source is near the interaural axis, the change in IID can span 15-30 JNDs as distance moves from 1 m to 0.12 m, providing a potentially strong binaural distance cue in the near field. Note that significant IIDs at low frequencies occur exclusively in the near field. In the far field, the low-frequency IID is small at all source directions.
2. In contrast to the IID, the ITD is roughly independent of distance in the near field.
3. Both the sphere model and the KEMAR measurements indicate that the average low-frequency gain of the HRTF increases more rapidly than the average high-frequency gain as the source approaches the head. Thus, the sound reaching the ears is effectively low-pass filtered as the source approaches the head. This filtering may serve as a spectral distance cue in the near field.
4. The HRTF at the contralateral ear is dominated by a complex interference pattern from sound propagating around the head by different paths. Up to about 2 kHz, this effect causes an increase in amplitude in the HRTF (a "bright spot") when the ear is located directly opposite the source. At higher frequencies, the interference pattern results in a complex series of ridges and notches which change with frequency and azimuth. This interference pattern at the contralateral ear tends to dominate the detail of the IID at high frequencies.
5. The discrepancy between the orientation of the source relative to the ear and the orientation of the source relative to the head causes a remapping of the high-frequency azimuth cues at the ipsilateral ear. In general, the features of the transfer function tend to be shifted laterally as the source approaches the head.

6. HRTF measurements at three elevations and three distances indicate that the high-frequency features of the HRTF which vary systematically with elevation are not strongly dependent on distance. These data indicate that elevation localization may not be significantly different in the near and far fields.

In summary, it is clear that the distance-dependent attributes of the HRTF in the near field provide possible binaural distance cues which are unavailable for more distant sources. The changes in azimuth and elevation cues are less dramatic, and their effect on near-field localization is more difficult to predict. Two experiments are underway to measure localization accuracy in the near field and determine whether listeners are able to make use of the available distance cues in that region.

## 14.0 REFERENCES

- Begault, D. R. (1987). *Control of Auditory Distance*. Ph.D. thesis, University of California, San Diego.
- Belendiuk, K., & Butler, R. (1977). Spectral cues which influence monaural localization in the horizontal plane. *Perception and Psychophysics*, 22, 353-358.
- Blauert, J. (1983). *Spatial Hearing*. Cambridge, MA: MIT Press.
- Brungart, D. (1998). *Near-Field Auditory Localization*. Ph.D. thesis, Massachusetts Institute of Technology.
- Brungart, D., & Rabinowitz, W. (1996). Auditory localization in the near-field. In *Proceedings of the Third International Conference on Auditory Display*.
- Burton, P. (1990). Audio Speaker. United States Patent 4,924,506.
- Butler, R. (1987). An analysis of the monaural displacement of sound in space. *Perception and Psychophysics*, 41, 1-7.
- Butler, R., Humanski, R., & Musicant, A. (1990). Binaural and monaural localization of sound in two-dimensional space. *Perception*, 19, 241-256.
- Carlile, S., & Pralong, D. (1994). The location dependent nature of perceptually salient features of the head-related transfer function. *Journal of the Acoustical Society of America*, 95, 3445-3459.
- Chan, J., & Geisler, C. (1990). Estimation of eardrum acoustic pressure and of ear canal length from remote points in the ear canal. *Journal of the Acoustical Society of America*, 87, 1237-1247.
- Coleman, P. (1963). An analysis of cues to auditory depth perception in free space. *Psychological Bulletin*, 60, 302-315.
- Duda, R. O., & Martens, W. L. (1997). Range dependence of the HRTF for a spherical head. In *Proceedings of the IEEE ASSP Workshop*.

- Firestone, F. (1930). The phase difference and amplitude ratio at the ears due to a source of pure tone. *Journal of the Acoustical Society of America*, 2, 260–270.
- Foster, S. (1986). Impulse response measurements using Golay codes. *IEEE Acoustics, Speech and Signal Processing*, 2, 929–932.
- Gardner, W., & Martin, K. (1995). HRTF measurements of a KEMAR. *Journal of the Acoustical Society of America*, 97, 3907–3908.
- Greene, D. (1968). Comments on 'Perception of the range of a sound source of unknown strength'. *Journal of the Acoustical Society of America*, 44, 634.
- Hartley, R., & Frey, T. (1921). The binaural location of pure tones. *Physical Review*, 18, 431–442.
- Hershkowitz, R., & Durlach, N. (1969). Interaural time and amplitude JNDs for a 500-Hz tone. *Journal of the Acoustical Society of America*, 46, 1464–1467.
- Hirsch, H. (1968). Perception of the range of a sound source of unknown strength. *Journal of the Acoustical Society of America*, 43, 373–374.
- Kinsler, L., & Frey, A. (62). *Fundamentals of Acoustics*. New York: Wiley & Sons.
- Kuhn, G. (1979). The pressure transformation from a diffuse sound field to the external ear and to the body and head surface. *Journal of the Acoustical Society of America*, 65, 991–1000.
- McKinley, R., Ericson, M., & D'Angelo, W. (1994). Three dimensional audio displays: development, applications, and performance. *Aviation, Space, and Environmental Medicine*, 65, A31–38.
- Mehrgardt, S., & Mellert, V. (1977). Transformation characteristics of the external human ear. *Journal of the Acoustical Society of America*, 61(6), 1567–1576.
- Middlebrooks, J., & Green, D. (1991). Sound localization by human listeners. *Annual Review of Psychology*, 42, 135–139.
- Middlebrooks, J., Makous, J. C., & Green, D. M. (1989). Directional sensitivity of sound-pressure levels in the human ear canal. *Journal of the Acoustical Society of America*, 86, 89–107.

- Molino, J. (1970). Perceiving the range of a sound source when the direction is known. *Journal of the Acoustical Society of America*, 53, 1301-1304.
- Moller, H., Sorensen, M., Hammershoi, D., & Jensen, C. B. (1995). Head-related transfer functions of human subjects. *Journal of the Audio Engineering Society*, 43, 300-320.
- Musicant, A. D., & Butler, R. A. (1984). The psychophysical basis of monaural localization. *Hearing Research*, 14, 185-190.
- Oldfield, S., & Parker, S. (1984). Acuity of sound localization: a topography of auditory space. II: pinna cues absent. *Perception*, 13, 601-617.
- Oldfield, S., & Parker, S. (1986). Acuity of sound localization: a topography of auditory space. III: Monaural hearing conditions. *Perception*, 15, 67-81.
- Pralong, D., & Carlile, S. (1994). Measuring the human head related transfer functions: A novel method for the construction and calibration of a miniature 'in ear' recording system. *Journal of the Acoustical Society of America*, 95, 3435-3444.
- Rabinowitz, W., Maxwell, J., Shao, Y., & Wei, M. (1993). Sound localization cues for a magnified head: implications from sound diffraction about a rigid sphere. *Presence*, 2, 125-129.
- Rayleigh, L. (1907). On our perception of sound direction. *Philos. Mag.*, 13, 214-232.
- Rogers, M., & Butler, R. (1992). The linkage between stimulus frequency and covert peak areas as it relates to monaural localization. *Perception and Psychophysics*, 52, 536-546.
- Shaw, E. A. G., & Teranishi, R. (1968). Sound pressure generated in an external-ear replica and real human ears by a nearby point source. *Journal of the Acoustical Society of America*, 44, 240-249.
- Shaw, E. (1974). Transformation of sound pressure level from the free field to the eardrum in the horizontal plane. *Journal of the Acoustical Society of America*, 56, 1848-1861.
- Slattery, W., & Middlebrooks, J. (1994). Monaural sound localization: acute versus chronic unilateral impairment. *Hearing Research*, 75, 38-46.
- Stewart, G. (1911). The acoustic shadow of a rigid sphere with certain applications in architectural acoustics and audition. *Physical Review*, 33, 467-479.

- v. Bekesy, G. (1960). *Experiments in Hearing*. New York: Wiley.
- Wallach, H. (1939). On sound localization. *Journal of the Acoustical Society of America*, 10, 270-274.
- Wenzel, E. (1991). Localization in virtual acoustic displays. *Presence*, 1, 80-107.
- Wightman, F., & Kistler, D. (1989). Headphone simulation of free-field listening. II: Psychological validation. *Journal of the Acoustical Society of America*, 85, 868-878.
- Wightman, F., & Kistler, D. (1992). The dominant role of low-frequency interaural time differences in sound localization. *Journal of the Acoustical Society of America*, 91, 1648-1660.

# APPENDIX A: ACOUSTIC POINT SOURCE FOR NEAR-FIELD HRTF MEASUREMENTS

## A.1 Abstract

An approximation to an acoustic point source has been developed which produces relatively non-directional acoustic signals over a wide frequency range (200 Hz-15 kHz). The invention differs from previous point-source systems in several important ways: 1) The use of a high-output electro-dynamic horn driver in place of a conventional cone loudspeaker to power the unit; 2) The use of a relatively long, flexible tube to carry the signal away from the driver, allowing the driver unit to be acoustically isolated from the point source and also allowing easy placement of the point source; 3) The use of a rigid sleeve around the distal end of the tube to allow more convenient placement of the source; 4) The use of an electromagnetic tracking system to accurately measure the effective location of the point source without interfering with its output. The resulting invention has a variety of potential applications where a compact, non-directional, high-output source is required, including acoustic and psycho-acoustic measurements in the near field.

## A.2 Purpose

Under certain circumstances, it is desirable to make acoustic measurements with an acoustic "point source". A point source is defined as an infinitesimally small sound source which produces a finite quantity of acoustic power. Usually it is modeled as a pulsating sphere of negligible dimensions producing a finite volume velocity at its surface.

Point sources have two important characteristics which cannot be duplicated in any physically realizable acoustic transducers. They radiate sound from a single location in space, and they radiate sound omnidirectionally. Unfortunately, it is impossible to build an infinitesimally small sound transducer with these characteristics. While it is, of course, possible to build a small loudspeaker, there is a clear tradeoff in loudspeaker design between small size and low-frequency output. The invention described here is unique in that it is able

to generate sound from a compact region of space which is both largely non-directional at relatively high frequencies and relatively powerful at low frequencies. The current rendition of this system generates sound from a location only 1.3 cm in diameter, is capable of generating reasonably strong output from 200 Hz to 15 kHz, and has a 3 dB beamwidth of approximately  $120^\circ$  at 15 kHz. The invention is also equipped with an electromagnetic position sensing system that allows accurate measurement of the effective position of the source.

The purpose of the system described herein is to enhance the accuracy of acoustic measurements in situations where conventional loudspeakers capable of producing enough low-frequency output are not sufficiently small or non-directional. An example of such an application is the measurement of Head-Related Transfer Functions (HRTFs) in the near field. The HRTF is the transfer function from the pressure at a sound source at some location in space to the pressure that actually reaches the eardrums of a human listener. This transfer function includes the propagation of sound from the source to the head, the diffraction of the head and torso, the spectral shaping of the outer ear or pinna, and the ear-canal resonance. Historically, most HRTF measurements have been made at distances of 1 m or more, where the dimensions of a loudspeaker are essentially negligible (they extend only over a few degrees in azimuth and elevation) and the location of the source relative to the head is easily determined. Current research efforts are underway to measure HRTFs at distances less than 1 m. At locations near the head even a relatively small loudspeaker can extend over a region  $25^\circ$  or more across, and the exact orientation of the source relative to the head is more difficult to determine. In order to measure the near-field HRTF at a well defined location, a point source with some mechanism for accurate positioning is required.

The system is also useful in applications other than measurements when a compact, wide-bandwidth, non-directional source is useful. For example, the point source can be used to conduct psycho-acoustic localization experiments in the near field.

### A.3 Background

Clearly the most conventional transducer for generating an acoustic signal from an electrical input is a loudspeaker. HRTF measurements, for example, have traditionally used conventional loudspeakers, 7 cm or larger in diameter, to generate the acoustic stimulus. At distances of 1 m or more, such loudspeakers are perfectly adequate. At close distances, however, there are serious problems associated with loudspeaker measurements:

- The precise location of a loudspeaker is not well defined in the near field. The stimulus is generated by the entire diaphragm of the loudspeaker, and at close distances this may extend over a large region of space: at 12 cm, for example, a 7 cm loudspeaker covers an arc in excess of 30°. The HRTF measured will be, in effect, the average HRTF over the entire region covered by the loudspeaker.
- The directional properties of the loudspeaker may taint the HRTF. When the speaker is near the listener, the high-frequency directionality of the speaker will cause the sound pressure reaching the head and torso to vary according to the orientation of that region relative to the speaker. This may significantly effect the measured HRTF.
- The axial response of a loudspeaker is complicated by its distributed geometry at very close distances. At distances less than  $2\frac{a^2}{\lambda}$ , where  $a$  is the radius of the loudspeaker and  $\lambda$  is the wavelength of the sound, the intensity along the axis of the loudspeaker does not decrease monotonically with distance, but rather passes through a series of maxima of constant amplitude with intervening nulls (Kinsler & Frey, 62). For a 15 kHz sound generated by a 7 cm loudspeaker, this effect complicates measurements at distances less than 10 cm from the surface of the head (approximately 20 cm from the center of the head).
- A loudspeaker is generally large enough to provide a reflective surface when sufficiently close to the head. Sound generated by the speaker might be reflected off the head, then be reflected again off the source and back toward the head. These second-order reflections could corrupt a near-field HRTF measurement.

For these reasons, an ordinary loudspeaker cannot be used effectively to make near-field HRTF measurements. The key to eliminating the problems associated with loudspeaker measurements is reducing the effective area of the source. Every realizable transducer has finite dimensions, and therefore generates a positive particle velocity over some finite region of space. The sound pressure generated by such a source at particular location in space is found by dividing the moving surface of the source into infinitesimal regions. The contribution of each region is determined by assuming that region is a point source with a certain volume velocity. The surface integral of these contributions over the area of the transducer determines the total signal. In acoustic measurements of the transfer function from a sound source at a particular location in space to a receiver at some other location in space, any measurement with a conventional transducer will in fact be the average transfer function

over the region covered by the transducer. In order to control the exact location of a sound source, it is necessary to make the area of the transducer as small as possible.

One possible approach to this problem is the use of extremely small loudspeakers. This would certainly reduce the problems of location, directionality, axial response, and reflections described above. However, due to radiation impedance, there is an inverse relation between the efficiency of a loudspeaker at low frequencies and the size of the loudspeaker. Thus extremely small loudspeakers cannot effectively reproduce wide-band stimuli.

A second approach to the problem is to generate a wide-band stimulus with a relatively large conventional cone loudspeaker and connect this speaker, through an enclosed cavity, to a small diameter metal tube. The sound then propagates down the tube and radiates from the small orifice at the opening of the tube. This is exactly the approach used by Shaw and Teranishi (Shaw & Teranishi, 1968). They connected a loudspeaker to small enclosure, which then opened into a rigid tube, 30 cm long and 1 cm in diameter. Sound propagated down the tube, and approximated a point source at its opening. A pressure microphone at the opening of the tube was used to actively control the output of the source and maintain a flat frequency response. This approach was effective, in that it produced output from 1 kHz to 15 kHz and had a 2 dB beamwidth of 90° at 15 kHz, but apparently was incapable of producing a stimulus below 1 kHz. There are two reasons why this type of system cannot generate low frequency sounds. First, the radiation impedance of the small tube is very high, especially at low frequencies, and a conventional cone loudspeaker is simply not powerful enough produce much output below 1 kHz. Second, it is extremely difficult to prevent low-frequency energy from leaking out of the loudspeaker enclosure. It generally takes extremely massive barriers to prevent the propagation of sound at low frequencies, and a point source enclosed with such massive baffling material would be unwieldy at best.

A third, relatively novel way to simulate an acoustic point source consists of a speaker with stretched, round membrane which is driven only at its center (Burton, 1990). If the membrane material is chosen carefully, vibrations propagate down the membrane at the same speed the sound waves propagate in air. This results in a hemispherically symmetrical sound radiation pattern. While this system approximates an acoustic point source, it still apparently requires a round membrane which may reflect scattered sound waves. Also, this system must be built from scratch and cannot be adapted from commercially available components.

The system described here, based on the second approach, significantly improves the

method used by Shaw and Teranishi in two ways. First, it uses a high-output electrodynamic horn driver in place of the loudspeaker. This driver is sufficiently powerful to drive even the high impedance of a small-diameter tube at low frequencies, and is more easily adapted down to the small diameter of the tube than a loudspeaker enclosure. Second, a long (3.5 m) section of flexible, thick-walled nylon tubing in place of the rigid metal tube used by Shaw and Teranishi. The use of flexible tubing has two important advantages. First, the driver can be located 1-2 m away from the opening of the tube. This allows the driver unit to be baffled with any amount of material to reduce leakage at low frequencies, and reduces the effect of any such leakage because the opening of the source is much closer to the receiver than the interfering leakage from the driver. Second, the flexible tube makes the actual placement of the source very convenient, and the actual source can easily be manipulated by hand without moving the massive driver unit. In fact, a specially shaped wand has been developed to allow a stationary operator to move the point source anywhere in the right hemisphere of a listener within 1 m of the head, which is particularly useful in near-field psychoacoustic measurements.

An additional feature of this system is the addition of an electromagnetic position sensor at the opening of the source. Electromagnetic sensors have been in use for a variety of applications for about 20 years, including sensing head and hand positions in virtual reality systems. The use of these systems for locating a source during an acoustic measurement has not been described previously. In HRTF measurements, source position has traditionally been controlled through the use of an automated source placement system, such as a revolving hoop with several speakers placed at regular intervals on the hoop (Wightman & Kistler, 1989). These systems are only able to control source location in azimuth and elevation.

In the near field, HRTFs are dependent on distance as well as direction, and some mechanism is required to allow accurate placement of the sound source in three dimensions. An electromagnetic tracking system is particularly well suited to this application, because it can measure both the orientation and location of the sensor. Obviously, the sensor cannot be at the exact location of the opening of the point source. It must be on the tube slightly back from the opening. However, if the end of the tube is mounted in a rigid sleeve, the combination of information about the orientation and XYZ coordinates of the source allow accurate measurement of the effective location of the opening of the point source. Note that the sensitivity of electromagnetic position sensors to metal and to magnetic fields preclude their use to measure the location of a loudspeaker, which includes a large magnet.

## A.4 Description of System

A diagram of the point source system is shown in Figure A-1. Basically, the system consists of four major components: a high-output acoustic driver, a long flexible plastic tube, a rigid plastic sleeve around the termination of the tube, and an electromagnetic position sensing system.

The point source is based on a high-performance, high-frequency acoustic driver, in this case an Electro-Voice DH1506 (1). This driver is designed for use in conjunction with a large exponential horn in high-output public address systems. When connected to such a horn, the driver is capable of generating extremely loud signals (104 dB SPL at 10 ft), and has a frequency range flat from 500 Hz to 3 kHz and with a controlled roll-off to 20 kHz. In this application, the driver is not connected to a horn but rather is connected directly to a length of 1.3 cm i.d. tubing. The opening of the driver is 3.5 cm in diameter, so a series of fittings are required to mate the tubing to the driver. First, a 1.5" (3.8 cm) to 1" (2.5 cm) copper fitting (2) is mounted over the threaded opening of the driver. Teflon pipe-fitting tape was used to fill the gap between the threaded driver opening and the smooth-walled copper fitting. A 1" (2.5 cm) to 0.75" (1.9 cm) brass brushing (3) fits into the copper fitting, and is connected directly to a 0.75" to 0.5" hose fitting (4), which acts as a right angle adapter. The driver assembly generates some sound due to leakage from the back of the driver, especially at 8-9 kHz. Therefore the entire assembly is wrapped in sound absorbing material (foam, blankets, etc.) to prevent this sound from propagating. The entire driver unit is quite heavy (> 5 kg in weight), primarily because of the large magnet used in the driver unit.

The tubing used in the point source is Tygon transparent tubing, with an internal diameter of 1.3 cm (0.5") and a wall thickness of 0.3 cm (0.125") (5). The tube is 3.5 m in length. The first 2.3 m of the tube are exposed openly. The next 1.16 m of the tube are encased in a sleeve constructed of PVC pipe with an internal diameter of 2.5 cm (1") (6). This sleeve, which acts as a placement wand, includes four straight lengths of pipe, 64 cm, 18 cm, 18 cm, and 15 cm long, and three 45° elbow joints.

The Tygon tubing extends 4 cm beyond the end of the PVC sleeve. At the opening, foam material fills the gap between the tubing and the interior of the sleeve. At the end of the tube, a small amount of acoustic foam has been forced into the opening to act as a terminating impedance. The amount of material used was adjusted to minimize resonances inside the tube (the quarter-wavelength resonance is at approximately 25 Hz).

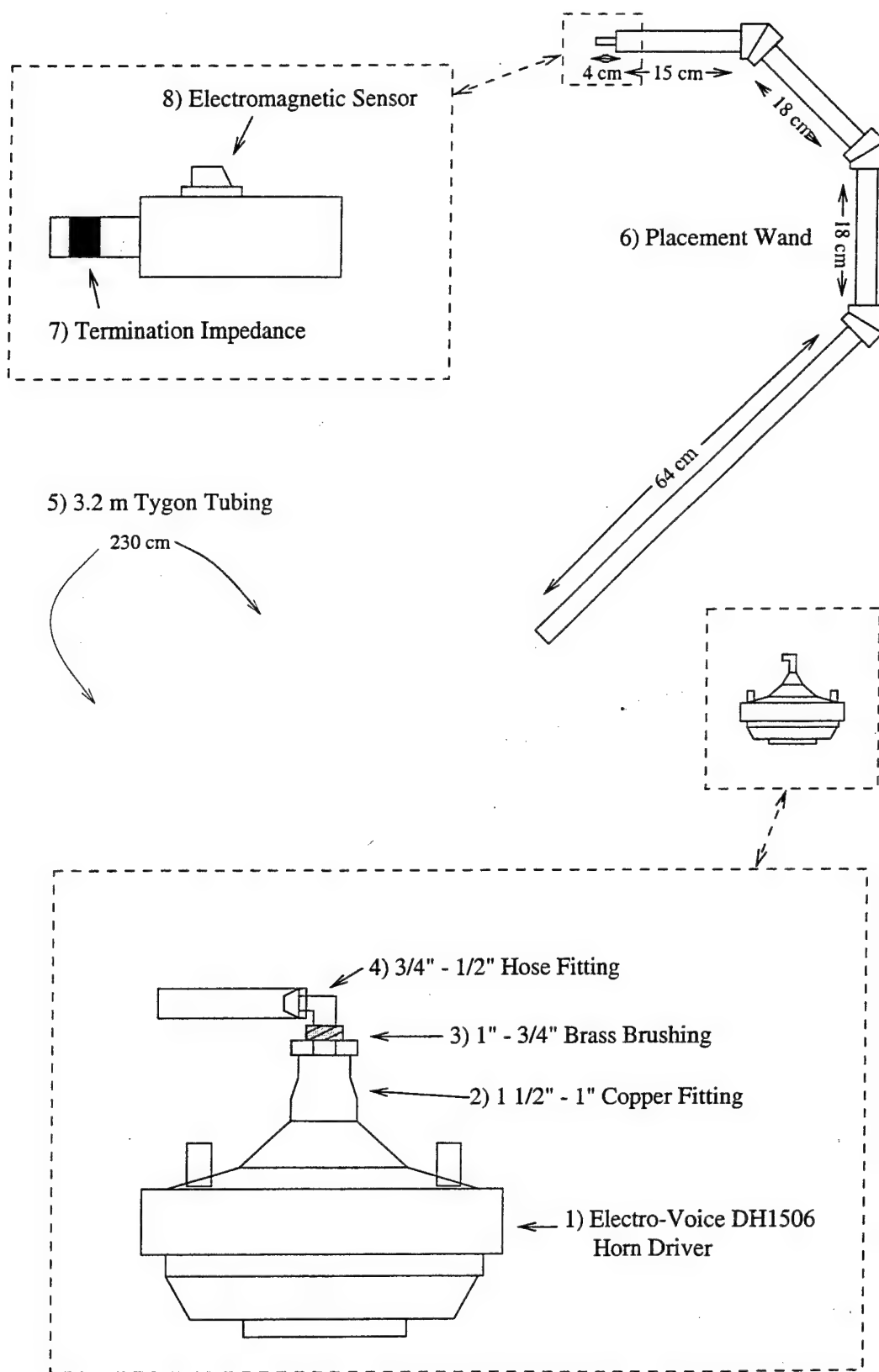


Figure A-1: Diagram of acoustic point source

Just before the end of the PVC sleeve, an electromagnetic sensor is attached by plastic cable ties. In this case, the sensor is from a Polhemus Electronics 3-Space Tracker, which is capable of determining the location of the sensor (relative to a separate electromagnetic source) within 0.25 cm in X,Y,Z coordinates, and the orientation of the sensor within 0.1° in roll, pitch, and yaw. The sensor is positioned so it remains in a fixed location relative to the opening of the tube (which is the effective location of the point source). Specifically, the center of the sensor is located 4 cm above and 6 cm behind the opening of the tube. The position of the opening of the tube can be found from the XYZ and roll (rl), azimuth (az), and elevation (el) coordinates produced by the 3-Space Tracker with the following equations:

$$x_{opening} = x_{sensor} + 6 \cos(az) \cos(el) + 4(\cos(az) \sin(el) \cos(rl) + \sin(az) * \sin(rl)), \quad (1)$$

$$y_{opening} = y_{sensor} + 6 \sin(az) \cos(el) + 4(\sin(az) \sin(el) \cos(rl) - \cos(az) * \sin(rl)), \quad (2)$$

$$z_{opening} = z_{sensor} - 6 \sin(el) + 4 \cos(el) \cos(rl). \quad (3)$$

## A.5 Characteristics of Point Source

Due to the unconventional transmission path from the driver to the opening of the tube, the frequency response of the system is quite erratic (Figure A-2). The response slopes gently upward from 200 Hz to 1 kHz, then goes through four local maxima up to 6 kHz. Above 6 kHz, the frequency response drops suddenly by 30 dB, and stays at this lower level (with several more local maxima) until dropping dramatically again at 15 kHz. The transfer function is, however, stable to changes in the configuration of the Tygon tubing, so the source can be moved without changing the response characteristics.

The non-directional response of the point source is shown in Figure A-3. As would be expected, the high-frequency sound radiated by the source drops off as the source is rotated away from normal incidence. The 3 dB beam-width of the source is about 120° at frequencies up to 15 kHz. At angles greater than 60°, the high-frequency response degrades rapidly, but the source appears to be completely omnidirectional at frequencies up to 2 kHz. In most

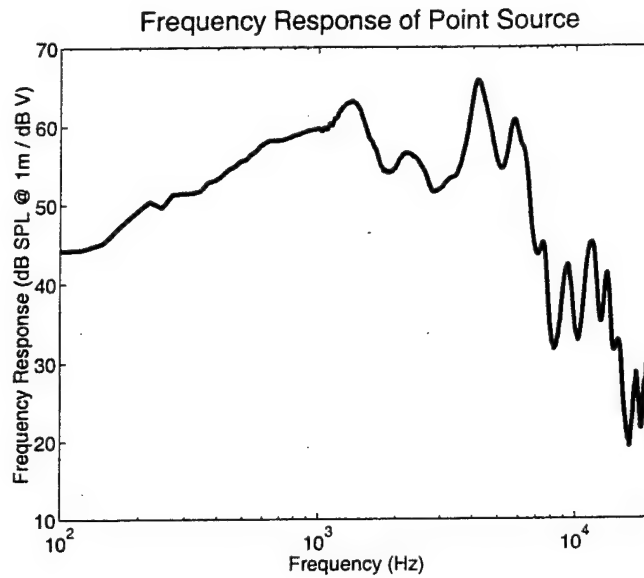


Figure A-2: Source transfer function. This figure shows the transfer function of the point source. The measurements were made in an anechoic chamber using a periodic chirp stimulus and a 1024-point FFT. One-third octave smoothing has been applied.

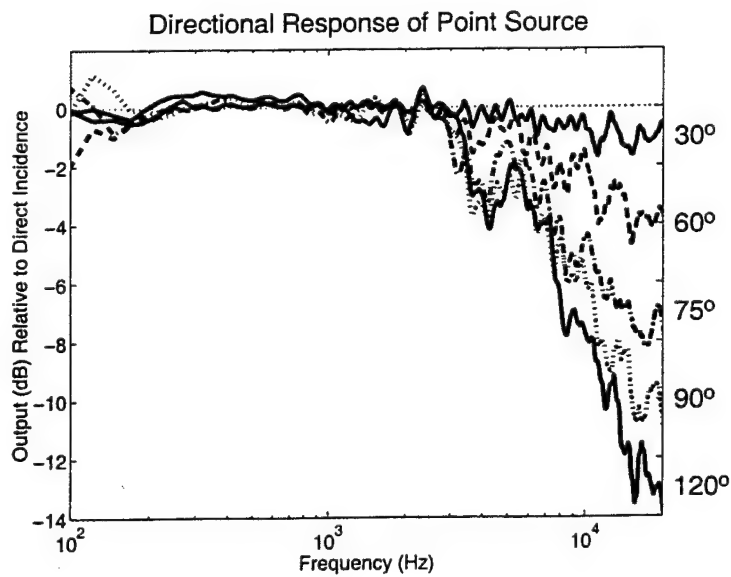


Figure A-3: Source directionality. The plots show the frequency response of the source at five source directions relative to normal incidence. The measurements were made in an anechoic chamber using a periodic chirp stimulus and a 1024-point FFT. One-third octave smoothing has been used.

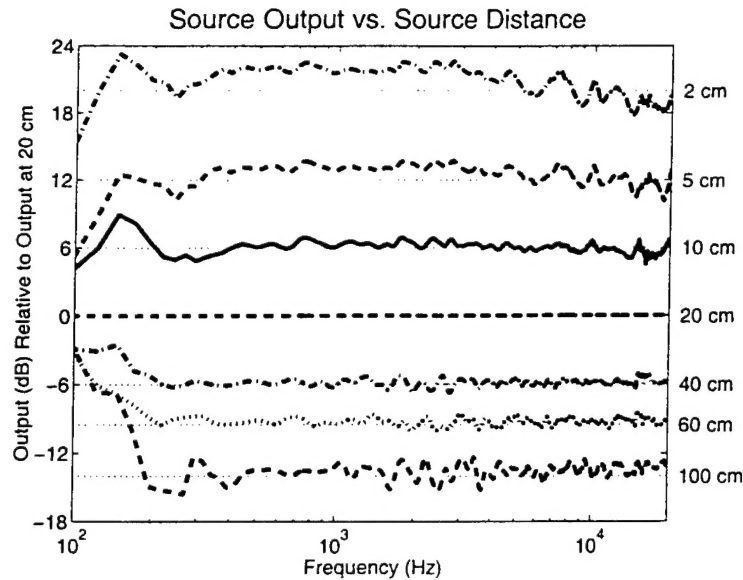


Figure A-4: Source response vs. distance in anechoic chamber. This figure shows the changes in the response of the point source at 6 different distances from the source tip, normalized to the response at 20 cm. The dotted lines represent the predicted source level for an acoustic point source. The measurements were made in an anechoic chamber using a periodic chirp stimulus and a 1024-point FFT. One-twelfth octave smoothing has been used. Note the discrepancies below 200 Hz are a result of the noise floor of the measurement.

practical applications, it should be possible to orient the source in such a way that the beamwidth of  $120^\circ$  covers the entire region under investigation.

In addition to being non-directional, a point source should generate a pressure wave that is inversely proportional to distance at all distances and all frequencies. Figure A-4 shows that the invention exhibits this behavior, except for a slight boost at low frequencies (less than 2 dB) when the source is within 2 cm of the receiver. This behavior indicates that the effective area of the source is quite small and that there are no significant leakage from the driver unit interfering with the measurements.

Another concern about the source is the possibility of non-linear operation from over-driving the compression driver unit. Figure A-5 shows the change in the measured output of the source at 20 cm when the input is attenuated. This system appears to be quite linear.

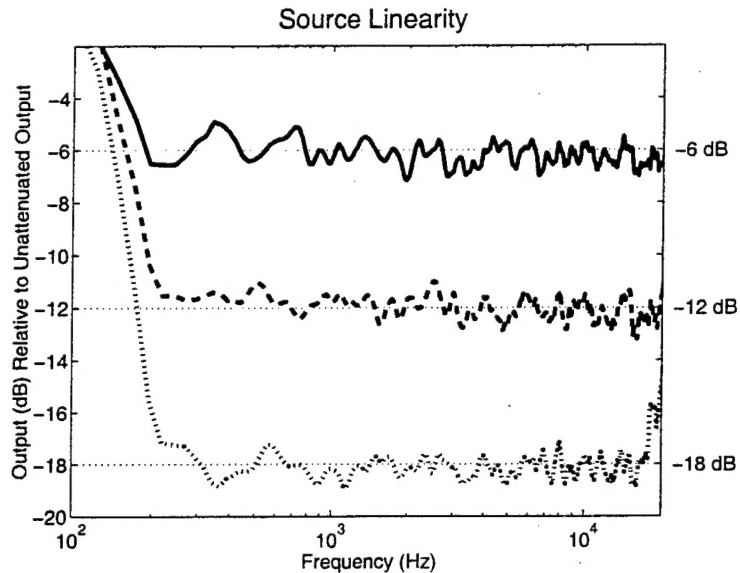


Figure A-5: Source linearity. This figure shows the change in the output of the source when the input is attenuated. The measurements were made in an anechoic chamber using a noise stimulus and a 1024-point FFT (hanning window). One-twelfth octave smoothing has been used. Errors below 200 Hz are a result of the noise floor.

## A.6 Use of the Point Source

In ordinary use, the point source driver would be connected directly to a conventional high-power audio amplifier, and driven by any reasonable signal generator. When making acoustic measurements, the source of the electromagnetic tracker would be placed in some fixed location relative to the system under test, and the XYZ location of the source relative to the system could be calculated directly with the previously described equations. In general, the placement wand would be clamped in place with a stand, and the sensor measurements would be used to move the source to the desired location relative to the system.

When rapid source placement is required (such as in a psychoacoustic experiment) the curvature of the placement wand is designed to allow a human operator to stand in a fixed location approximately 1 m away from a receiver (a human subject, for example), and be able to move the source to any location within 1 m of the receiver in the hemisphere closest to the operator. The 135° bend in the placement wand enables the operator to keep the source oriented in the direction of the receiver throughout this area, eliminating any undesirable effects due to source directionality at high frequencies. The electromagnetic sensor allows

a control computer to determine the exact location of the sound source rapidly even when manually placed by a human operator.

The irregular frequency response of the transducer (Figure A-2) limits the use of the point source directly in many applications, but it is generally possible to compensate for the irregular response. In acoustic measurements, the stability of this transfer function allows it to be completely removed from a measurement. For example, in measuring Head-Related Transfer Functions, the desired quantity under test is the ratio of sound pressure at the eardrum to the free-field sound pressure at the center of the head. In this application, the point source can be used directly because its transfer function is present in both measurements and is eliminated when the ratio between the measurements is calculated.

In other applications, an audio signal with a relatively flat frequency spectrum is required. In this case, the input signal to the point source can be electronically filtered by the inverse of its frequency response. This technique can be used to flatten the spectrum of the audio signal generated by the source. Figure A-6 shows the output of the point source when this technique was used to generate a low-pass filtered noise signal (6dB/octave rolloff above 200 Hz). Despite the large peaks and notches in the frequency response of the system (Figure A-2), many of which are 20 dB or larger in magnitude, the flattened response is within 1-2 dB of the desired response at all frequencies. Here a low-pass filtered signal was chosen because the response characteristics of the system allow greater non-distorted output with a low-pass filtered signal than with a flat signal. A white-noise spectrum could also be generated, but only a lower total output could be achieved.

A final note is in order about the use of the point source. The length and complexity of the propagation path from the source to the opening of the tube and consequent reflections result in a long and complicated impulse response, as well as a tendency towards intermodal distortion in high-output narrow-band signals. For these reasons, the source is better suited to broadband stimuli, and especially to noise signals, than to narrow-band or speech signals. Also, it is better suited to measurements using spectral averaging than to measurements which attempt to evaluate the impulse system directly.

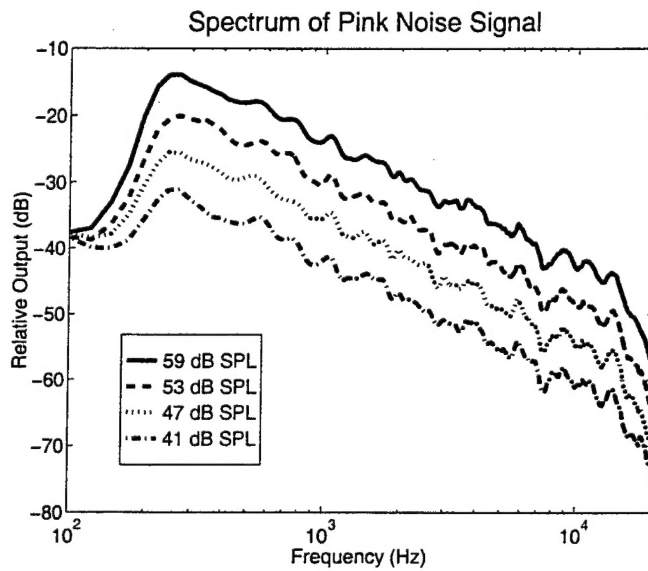


Figure A-6: Flattened Pink Noise Output. This figure shows the electronically flattened pink noise stimulus to be used in the proposed experiment. Four output levels (at 1 m) are shown. The measurements were made in an anechoic chamber using a noise stimulus and a 1024-point FFT (hanning window). One-twelfth octave smoothing has been used.